ENGINEERING

USCIPI Report 680

(1)

UNIVERSITY OF SOUTHERN CALIFORNIA

IMAGE RESTORATION BY SPLINE FUNCTIONS

by

Mohammad Javad Peyrovian

August 1976

Image Processing Institute
University of Southern California
University Park
Los Angeles, California 90007

Sponsored by
Advanced Research Projects Agency
Contract No. F-33615-76-C-1203
ARPA Order No. 3119

IPI

IMAGE PROCESSING INSTITUTE

USCIPI Report 680

IMAGE RESTORATION BY SPLINE FUNCTIONS

by

Mohammad Javad Peyrovian

August 1976

Image Processing Institute
University of Southern California
University Park
Los Angeles, California 90007

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Avionics Laboratory under Contract No. F-33615-76-C-1203, ARPA Order No. 3119.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Image Processing Institute University of Southern California, Los Angeles, CA	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

IMAGE RESTORATION BY SPLINE FUNCTIONS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, August 1976

5. AUTHOR(S) (First name, middle initial, last name)

Mohammad J. Peyrovian

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
31 August 1976	121	64

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
F-33615-76-C-1203 b. PROJECT NO.	USCIPI Report 680
c. ARPA Order No. 3119 d.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

PUBLIC

Approved for release: distribution unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

Spline functions, because of their highly desirable interpolating and approximating characteristics, are used as a potential alternative to the conventional pulse approximation method in digital image processing. For uniformly spaced knots, a class of spline functions called B-splines has the useful properties of shift invariance, positiveness, and convolutional and local basis properties. These properties are exploited in image processing for linear incoherent imaging systems.

The problem of image degradation in a linear imaging system is described by a superposition integral. For simulation of degradation and restoration by means of a digital computer, the continuous imaging model must be discretized. Thus, a theoretical and experimental study of quadrature formulae, particularly monospline and best quadrature formulae in the sense of Sard, is presented. It is shown that a good choice of degree for a monospline highly depends on the frequency content of the integrand, and in most cases, a cubic monospline generates less error than the pulse approximation method and Newton-Cotes quadrature formulae.

In space-invariant imaging systems, the object and point-spread function are represented by B-splines of degrees m and n. Exploiting the convolutional property, the deterministic part of the blurred image is a spline function of degree m + n + 1. A minimum norm principle leading to pseudo-inversion is used for the restoration of space-variant degradations and underdetermined and overdetermined models. Space-variant point-spread functions that describe astigmatism and

DD FORM 1 NOV 65 1473

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

curvature-of-field are derived and coordinate transformations are applied to reduce the dimensionality. The singular-value-decomposition technique is used for solution of the simplified equations.

For noisy blurred images, a controllable smoothing criteria based on the locally variable statistics and minimization of the second derivative is defined, and the corresponding filter, applicable to both space-variant and space-invariant degradations, is obtained. The parameters of the filter determine the local smoothing window and overall extent of smoothing, and thus the trade-off between resolution and smoothing is controllable in a spatially nonstationary manner. Since the matrices of this filter are banded circulant or Toeplitz, efficient algorithms are used for matrix manipulations.

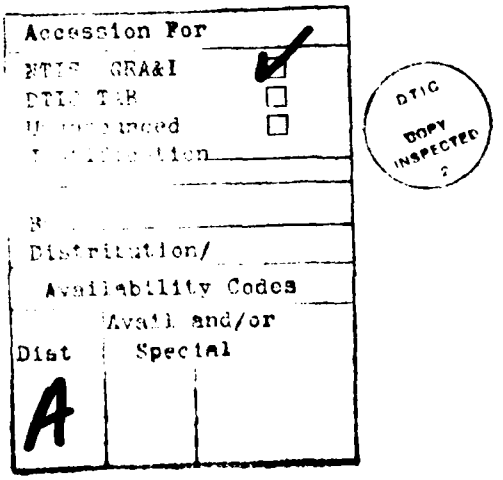
* *

14. Key Words: Digital Image Processing, Monospline Quadrature Formulae, Image Restoration, Spline Functions, Smoothing Splines.

ACKNOWLEDGEMENT

The author wishes to express his sincere gratitude to the chairman of his committee, Professor Alexander A. Sawchuk, for his guidance throughout the course of this research. Without Professor Sawchuk's suggestions and encouragement this dissertation could not have been possible. The author also feels deep appreciation towards Professors Harry C. Andrews and John G. Pierce, the other members of his committee, for their assistance and contributions.

This research was supported by the Advanced Research Projects Agency of the Department of Defense monitored by the Air Force Avionics Laboratory, Wright-Patterson Air Force Base.



ABSTRACT

Spline functions, because of their highly desirable interpolating and approximating characteristics, are used as a potential alternative to the conventional pulse approximation method in digital image processing. For uniformly spaced knots, a class of spline functions called B-splines has the useful properties of shift invariance, positiveness, and convolutional and local basis properties. These properties are exploited in image processing for linear incoherent imaging systems.

The problem of image degradation in a linear imaging system is described by a superposition integral. For simulation of degradation and restoration by means of a digital computer, the continuous imaging model must be discretized. Thus, a theoretical and experimental study of quadrature formulae, particularly monospline and best quadrature formulae in the sense of Sard, is presented. It is shown that a good choice of degree for a monospline highly depends on the frequency content of the integrand, and in most cases, a cubic monospline generates less error than the pulse approximation method and Newton-Cotes quadrature formulae.

In space-invariant imaging systems, the object and point-spread function are represented by B-splines of degrees m and n. Exploi-

ting the convolutional property, the deterministic part of the blurred image is a spline function of degree m+n+1. A minimum norm principle leading to pseudo-inversion is used for the restoration of space-variant degradations and underdetermined and overdetermined models. Space-variant point-spread functions that describe astigmatism and curvature-of-field are derived and coordinate transformations are applied to reduce the dimensionality. The singular-value-decomposition technique is used for solution of the simplified equations.

For noisy blurred images, a controllable smoothing criteria based on the locally variable statistics and minimization of the second derivative is defined, and the corresponding filter, applicable to both space-variant and space-invariant degradations, is obtained. The parameters of the filter determine the local smoothing window and overall extent of smoothing, and thus the trade-off between resolution and smoothing is controllable in a spatially non-stationary manner. Since the matrices of this filter are banded circulant or Toeplitz, efficient algorithms are used for matrix manipulations.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

LIST OF FIGURES

LIST OF FIGURES (Cont'd)

Chapter 1

INTRODUCTION

The objective of image restoration is the reconstruction of a recorded image towards an ideal object by inversion of the degrading phenomena. These phenomena include such imperfect imaging circumstances as defocus, motion blur, optical aberrations, and noise [1], [2]. The pioneers of this field in the modern sense were Marechal and his co-workers [3] who recognized in the 1950's the potential of optical spatial filtering for restoring blurred photographs. Their success stimulated others to study image restoration by optical compensation of the degradations. However, it was the versatility of digital computers and the space program of the sixties with its need for high quality imagery that provided the necessary means and motivation for the development of the field. With digital processing it is possible to overcome many limitations of optical filtering and to explore new approaches which have no conceivable optical counterparts.

Restoration techniques require some knowledge concerning the degradation phenomena, and this knowledge may come from an analytical model, statistical model, or other a priori information of the imaging system. Thus considerable emphasis must be placed on the sources and models of degradation. In general, an exact

degradation model is too complicated to be used. However, for many cases of practical interest, a quite accurate model is given by a linear smoothing operation due to the optical imperfection followed by the addition of noise [4].

The earlier restoration techniques, mostly optically oriented, attempted to remove the degradation by inverse filtering [5]. Using the Fourier transform properties of lenses, the Fourier transform of the degraded image is simply multiplied by the inverse of the Fourier transform of the blurring function. This method is not without limitations and shortcomings. First, in many practical cases such as motion blur and defocusing, the Fourier transform of the blur function has zeros at spatial frequencies within the range of interest, and since the inverse of zero is undefined, the method breaks down for such cases. Second, for noisy images, this method enhances the high frequency component of the noise. Various modifications have been suggested to overcome these drawbacks, but all of them are ad hoc and intuitive [3], [5], [6], [7], [8]. In spite of all the limitations, inverse filtering can yield reasonably good results where noise is not the limiting degradation.

The minimum mean-squared error (MSE) criterion has been used as an objective criteria for restoration of noisy images. Assuming the object and noise to be uncorrelated random processes with a known blur function, Helstrom [9] has proposed a filter for

image restoration based on minimum MSE principle. This filter is the same as the classical Wiener filter which was developed in the 1940's in the field of signal processing. For an unknown blur function, Slepian [10] has solved the same problem assuming the blur function to be a random process. Utilizing the transform properties of imaging systems, Pratt [11] has introduced generalized Wiener filtering with improved computational efficiency. Habibi [12] has shown that a lower triangular transformation yields an efficient suboptimal Wiener filter. The Fourier transform properties of the circulant matrices has been used to develop a computationally fast algorithm for solving the Wiener filter [13].

The Wiener filter has limitations and shortcomings. The minimum MSE principle, which is the objective criteria of a Wiener filter, is suspect in image restoration. It is well known that the human visual system demands a more accurate reproduction of regions where the intensity changes rapidly than of the regions with little change, and the sensitivity of the eye to a given error in intensity depends strongly upon the intensity [4]. The minimum MSE weights errors equally regardless of the intensity and its gradient. Moreover, a Wiener filter requires extensive a priori information, namely, the blur function and detailed knowledge of object and noise covariance functions. Finally, since the Wiener filter is derived by the Fourier transform properties of space-invariant degradations and

stationary assumptions of the object and noise, the filter is not applicable to space-variant degradations and non-stationary objects

Constrained restoration has been introduced as an alternative to overcome some short-comings of the Wiener filter. Hunt [14] has proposed a constrained least square filter, in which by judicious choice of some variables one can minimize higher order derivatives, eye model effects, or even achieve the Wiener filter. Stockham and Cole [15] have suggested a geometrical mean filter between the inverse filter and Wiener filter. Utilizing linear equality and inequality constraints has led to constrained restoration techniques [16]. The non-negative nature of image intensity has been the leading factor in some restoration techniques [17]. For unknown blur functions, the concept of homomorphic systems [18] has been employed to estimate the point-spread function from the degraded image by taking averages of image segments in the log-spectral domain [19]. A detailed comparison of these restoration techniques is given by Hunt [20].

For space-variant degradations, the problem of image restoration is much more difficult because Fourier techniques cannot be used. Generally, there are twice as many independent variables in a space-variant system as in a space-invariant system, and this increased dimensionality is the major analytical and computational difficulty. A method, analogous to Fourier techniques, has been

presented in terms of the degrading system eigenfunction [21], however, it is not known how to find a complete set of eigenfunctions or even if they exist. Sawchuk [22] has shown that for certain space-variant systems the degradation can be transformed to be space-invariant by an appropriate selection of coordinates.

The following is an outline of this dissertation and a summary of the contributions.

In Chapter 2 background on the problem of image degradation and restoration in a continuous model is discussed. The mathematical representation of this model, inverse filtering and the Wiener filter are studied briefly.

Chapter 3 is devoted to discrete representations of the continuous model for implementation on a digital computer. The pulse approximation method has been the simplest and the most common method in digital image processing, however, the accuracy of this technique is suspect in image discretization. It is shown that numerical analysis techniques, particularly monospline quadrature formulae, lead to a more accurate discrete model. The results are compared with extreme cases, namely, the pulse approximation method and the Newton-Cotes quadrature formulae. B-splines, because of their desirable characteristics and the useful properties of shift invariance, positiveness, and their convolutional and local basis properties, are studied and suggested for discrete representation of the continuous

model.

The restoration of noiseless images are presented in Chapter 4. For space-invariant imaging systems, the object and point-spread function are represented by B-splines of degrees m and n. The degree of B-spline must be selected with respect to the continuity and frequency content of the approximated function. Exploiting the convolutional property, the blurred image is a B-spline of degree m+n+1. It is shown that B-spline produces a better quality restoration than the conventional pulse approximation method. Pseudo-inversion based on the minimum norm principle is used for the restoration of space-variant degradations, overdetermined models and underdetermined models. With a linear incoherent system, the space-variant point-spread functions that describe imaging in the presence of astigmatism and curvature-of-field are derived and coordinate transformations are applied to reduce the dimensionality. The singular-value-decomposition techniques analogous to inverse Fourier filtering are used for pseudo-inverse solution of the simplified equations.

Image restoration by spline functions in the presence of noise is covered in Chapter 5. A controllable smoothing criteria based on the locally variable statistics and minimization of the second derivative is defined, and the corresponding filter, applicable to both space-invariant and space-variant degradations, is obtained. The

parameters of the filter determine the local smoothing window and overall extent of smoothing, and thus the tradeoff between resolution and smoothing is controllable in a spatially non-stationary manner. The interesting properties of this filter has made it capable of restoring signal-dependent noisy images, and it has been successfully applied for filtering images degraded by film-grain noise. Since the matrices of this filter are banded, circulant or Toeplitz, efficient algorithms are used for matrix manipulations.

Finally, conclusions and recommendations for further research are given in Chapter 6.

Chapter 2

IMAGE RESTORATION IN A CONTINUOUS MODEL

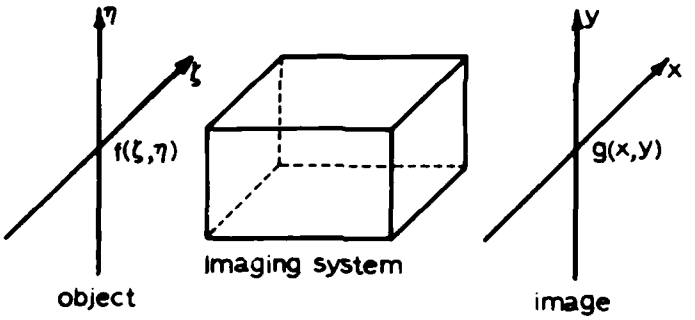
In this chapter, the problem of image degradation and restoration in a continuous model is discussed. Section 2.1 presents a mathematical model for a linear imaging system. In sections 2.2 and 2.3 respectively, restoration techniques for noiseless and noisy images are discussed.

2.1 Degradation in a Linear Imaging System

Let $g(x, y)$ be the image of an object $f(\xi, \eta)$ which has been degraded by the linear operator $h(x, y; \xi, \eta)$ such that

$$g(x, y) = \int_{-\infty}^{\infty}\int h(x, y; \xi, \eta) f(\xi, \eta) d\xi\, d\eta + n(x, y) . \qquad (2.1)$$

The first source of degradation, represented by $h(\cdot)$, is known as the impulse response or point spread function (PSF) of the imaging system. Physically, $h(\cdot)$ is assumed to be the image of a point source of light located at (ξ, η) in the object plane. The second source of degradation is an additive noise represented by $n(\cdot)$ which can only be characterized in statistical terms. Figure 2-1 represents a linear imaging system and the corresponding block diagram. Generally, the response $h(\cdot)$ in the image space varies with the position (ξ, η) of the input impulse and is called a space-variant point-spread function (SVPSF) in an optical context. If $h(\cdot)$ is isoplanatic, i.e., the form of $h(\cdot)$



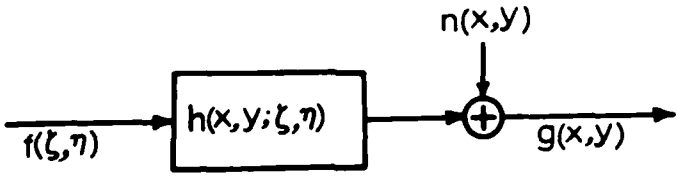


Figure 2-1. Linear imaging system model.

remains fixed in the image plane for all (ξ, η) positions, then the system is said to be spatially invariant and $h(\cdot)$ is called a space-invariant point-spread function (SIPSF). In this case, the impulse response is a function of two variables and the dimensionality of the system reduces considerably. The PSF $h(x, y; \xi, \eta)$ can then be written as $h(x-\xi, y-\eta)$ and the superposition integral (2.1) simplifies to a convolution.

$$g(x,y) = \int_{-\infty}^{\infty} h(x-\xi, y-\eta) f(\xi, \eta) d\xi d\eta + n(x,y). \qquad (2.2)$$

The mathematical representation given in (2.2) is general enough to cover many situations that occur in coherent and incoherent optical systems.

Some of the sources of degradation include: diffraction, motion degradation; defocusing and atmospheric turbulence. Diffraction in an optical imaging system is due to the limited aperture size and is an example of spatially invariant degradation. The blur function for a system with a circular aperture and incoherent illumination is given by [23]

$$h(x,y) = \left[\frac{J_1(2\pi\rho)}{\rho}\right]^2 \qquad (2.3)$$

where $\rho = (x^2+y^2)^{\frac{1}{2}}$ and J_1 is a Bessel function of the first kind, order one. Linear uniform motion degradation, defocusing and atmospheric turbulence are other examples of space-invariant degradations [4], [23]. In some cases, such as defocusing and

and atmospheric turbulence, the impulse response is separable, i.e., the function of two variables can be written as a product of two functions, each with one variable.

The assumption of space-invariance is not valid for certain degradations. Lens aberrations such as coma, astigmatism, curvature-of-field, motion blur where objects are at different distances from the camera and image plane tilt are examples of space-variant systems [26]. By an appropriate selection of coordinates, some of these degradations can be transformed into equivalent space-invariant [24], [25] systems.

The assumption of additive noise is broad enough to encompass different practical situations. Many of the noise sources (e.g., stray illumination, circuit noise, roundoff error) may be individually modeled as additive noise. Nevertheless, the assumptions of linearity and additive noise are subject to criticism because they are valid only over a certain dynamic range. The problem is that g is not directly available for processing. Instead, a nonlinear recording of g on a photographic emulsion is usually the only available measurement. It is possible to measure the nonlinear function to recover g over a larger dynamic range, but, any attempt at extending this range must ultimately be frustrated by a drastic increase in the noise level [4]. Also, the effect of film grain noise is far from being additive. Huang [27] has shown it could be modeled by a multiplicative process,

and more general signal-dependent models must be used to accurately describe the process [28].

After specifying assumptions and limitations, the next step is to clarify the necessary information. The model assumes that a complete knowledge of the impulse response h is available. This knowledge can be obtained analytically [23] or from edges or points in the image that are known to exist in the object [29]. As far as the noise is concerned, knowledge of the second order statistical properties is required. The noise is not necessarily white, but this assumption is often made.

Each restoration scheme given in the succeeding sections and chapters assumes some objective intuitively reasonable criteria of quality. Inverse filtering, for instance, attempts perfect resolution without regard to noise, while the Wiener filter minimizes the mean square error without regard to resolution. Although it is known that the human observer does not judge images according to mean square error [30], it has been found that reasonable results can be obtained by its use, especially for low contrast images. Moreover, mean square error leads to a very tractable mathematical structure, the regression model, which has been considerably explored in mathematical statistics.

2.2 Inverse Filtering

The idea of inverse filtering is very simple. Taking a Fourier

transform of both sides of the convolution expression (2.2) gives

$$G(u, v) = H(u, v)F(u, v) + N(u, v) \tag{2.4}$$

where u and v represent the x and y spatial frequencies and the upper-case letters represent Fourier transforms of the functions denoted by the corresponding lower-case letters. If the transfer function H(u, v) does not vanish at any point (u, v), the inverse filter R(u, v) is defined as

$$R(u, v) = 1/H(u, v) \ . \tag{2.5}$$

The restored image in the Fourier domain is obtained by multiplying both sides of (2.4) by R and taking an inverse Fourier transform. Thus

$$\hat{F}(u, v) = F(u, v) + N(u, v)/H(u, v) \tag{2.6}$$

$$\hat{f}(x, y) = f(x, y) + \mathcal{F}^{-1}(N/H) \tag{2.7}$$

where $\mathcal{F}^{-1}$ represents the inverse Fourier transform. Marechal et al. [3], Tsujiuchi [5], Harris [6], McGlamery [7], and Mueller and Reynolds [8] have applied this method with minor modifications. These modifications aim at two short-comings of the inverse filter. First, since H in most practical cases decreases rapidly for large values of u and v, and the noise has a fairly flat spectral density function, this filter amplifies the high frequency noise. Second, in many cases H has zeros at spatial frequencies within the range of

interest, and since division by zero is meaningless, the method breaks down. Therefore, a perfect restoration is impossible even in the noiseless case. All the modifications suggested to overcome these short-comings are intuitive and ad hoc. For example, Harris [6] suggests that the right hand side of (2.5) should be multiplied by a function $H_0(u, v)$ before taking the inverse Fourier transform. H_0 is chosen to be zero wherever the power spectral density of the signal is dominated by noise and H_0/H is chosen to be finite at the zeros of H. With this method, the problem of infinities is removed, but the restored image is a blurred version of $\hat{f}$ derived in (2.6) through the filter H_0. The requirements imposed on H_0 are intuitively reasonable. However, there are an infinite number of functions satisfying these requirements and one of them must be chosen arbitrarily.

It has been shown that with the modifications given in [3], [5]-[8], and in spite of the limitations and assumptions, inverse filtering can yield reasonably good restoration in many cases where noise is not dominant.

2.3 Wiener Filtering

The basic idea of Wiener filtering is to minimize the mean square error (MSE) between the original and restored images. MSE is an objective criteria for which the optimum restoration can be rigorously computed. The following is a brief outline of the mathematical derivation of the Wiener filter.

Assuming the object function f and the noise n be sample functions of different random processes, the minimum MSE estimate of f is $\hat{f}$ such that for each (x, y) of the object plane the error e given by

$$e = E[(f-\hat{f})^2] \tag{2.8}$$

is minimized, where E denotes expectation over the processes f and n. It can be proven that

$$\hat{f} = E(f|g) \tag{2.9}$$

in general.

Although (2.9) appears very simple, the computation of the conditional expectation is very difficult for arbitrary processes f and n. The problem becomes much simpler if we assume these processes to be stationary, independent, zero-mean Gaussian processes. In this case, it can be shown that the optimum nonlinear or linear estimate of (2.9) can be obtained by linear spatially invariant filtering of g. This minimization filter is easier to describe in the spatial frequency domain, and can be shown to be [4]

$$R = H^*/(HH^* + \varphi_n/\varphi_f) \tag{2.10}$$

where * denotes complex conjugate and φ_f and φ_n are power spectral densities of the processes f and n. With $\varphi_n = 0$, the Wiener filter becomes the inverse filter and when H = 0, R = 0 (unlike R = ∞ with the inverse filter). In order to determine R, both H and the noise-to-

signal ratio Φ_n/Φ_f must be known. In most cases Φ_n/Φ_f is assumed to be constant over the frequency range of interest.

For a better fidelity criterion than MSE, various modifications to the Wiener filter have been proposed. Hunt [14] has proposed a constrained least-square filter and Stockham and Cole [15] have suggested a filter which is the geometrical mean between the inverse filter and the Wiener filter. These filters have been implemented with very successful results.

Chapter 3

DISCRETIZATION OF THE CONTINUOUS MODEL

In the processing of images by digital computer, the continuous model of Eq. (2.1) must be discretized. In digital image processing, the information is necessarily finite and discrete in both amplitude and spatial position. Therefore, the continuous image field, and in most cases the impulse response, must be transformed into arrays of numbers. Generally, this transformation produces some error, i.e., the inverse transform of these arrays of numbers is not exactly the original image field.

Representing the continuous function by an array of samples, known as the pulse approximation method, is the simplest and the most common technique in image discretization. However, the accuracy of this technique is suspect in digital image modeling. Using numerical analysis methods, such as quadrature formulae, leads to a more accurate model. Spline functions, because of their highly desirable interpolating and approximating characteristics, are suggested as a potential alternative to the conventional pulse approximation method.

In this chapter, the problems of image sampling and quadrature formulae are analyzed. It is shown that spline functions are superior to both pulse approximation techniques and polynomials in discrete

representation of a continuous function and numerical solution of integral equations. Some experimental results are given in the last section of this chapter.

3.1 Pulse Approximation Method

The idea of the pulse approximation method is to represent a function $f(x)$ by an array of its sampled values taken on a countable set of points on the x axis. Clearly, if the sample points are close enough, the sampled data are an accurate representation of the original picture. Thus the function f can be reconstructed with sufficient accuracy by simple interpolation. Assuming Δx to be the distance between two subsequent points in a uniform sampling, the sampled function $f_s(x)$ is obtained by multiplying the original function by summation of δ functions as expressed by

$$f_s(x) = \sum_{i=-\infty}^{\infty} f(i\Delta x)\delta(x-i\Delta x) \quad . \tag{3.1}$$

Taking a Fourier transform of (3.1), the spectrum of the sampled function is given by [23]

$$F_s(u) = \sum_{i=-\infty}^{\infty} F(u - \frac{i}{\Delta x}) \tag{3.2}$$

where F is the Fourier transform of f and u is spatial frequency. Equation (3.2) indicates that the spectrum of the original function is infinitely repeated with a distance of $\frac{1}{\Delta x}$. Assuming the function f to be bandlimited, its spectrum F is non-zero over only a finite

interval R in the frequency domain. If Δx is sufficiently small, then the separation $\frac{1}{\Delta x}$ is large enough to assure that adjacent spectra do not overlap. If $2B_x$ represents the width of the rectangle that completely encloses the interval R, then non-overlapping is assured if

$$\Delta x \leq \frac{1}{2B_x} \quad . \tag{3.3}$$

Physically, this means that the function f must be sampled at a rate at least twice its highest frequency component or one-half the period of the finest detail within the function. If equality holds in Eq. (3.3), the function is said to be sampled at its Nyquist rate. If Δx is smaller or larger than this threshold, the function is oversampled or undersampled. With condition (3.3) satisfied, the exact reconstruction of the original function can be achieved by filtering the sampled data with an appropriate filter, for example a filter with a rectangular transfer function of width $2B_x$. In the spatial domain, the reconstruction operation in the spatial domain is

$$f(x) = \sum_{i=-\infty}^{\infty} f(\frac{i}{2B_x}) \operatorname{sinc}[2B_x(x-\frac{i}{2B_x})] \tag{3.4}$$

for a rectangular filter. Equation (3.4), known as the Whittaker-Shannon sampling theorem, indicates that the function is reconstructed exactly by an infinite sum of weighted sinc functions injected at each sample point.

3.2 Quadrature Formulae

The problem of image degradation, as stated in Eq. (2.1), is represented by an integral equation. In practical situations, the limits on this definite integral equation are not infinite. First, the degradation function h usually vanishes (or almost vanishes) beyond some point, and consequently, h is non-zero over a finite interval. Second, only a finite size of the object is of particular interest for restoration. With these considerations, the one-dimensional version of Eq. (2.1) is a definite integral

$$g(x) = \int_a^b h(x, \xi) f(\xi) d\xi \tag{3.5}$$

over a *finite interval* $[a, b]$.

To implement this continuous integral by a digital computer, a numerical technique, called a quadrature formulae (q.f.) must be employed. A q.f. is an approximation to a definite integral by a linear combination of values of the integrand, and perhaps also of some of its derivatives, at certain points of the interval of integration called the nodes of the q.f. [31]. A discrete version

$$g(x_i) = \sum_j c_{ij} h(x_i, \xi_j) f(\xi_j) \tag{3.6}$$

of Eq. (3.5) can be obtained by applying a q.f. Using vector space notation, the above equation simplifies to

$$\underline{g} = \underline{H}\,\underline{f} \tag{3.7}$$

where $\underline{H}$ is an M x N matrix with elements $h_{ij} = c_{ij} h(x_i, \tau_j)$. Assuming the coefficients $c_{ij} = 1$ is equivalent to the pulse approximation method. For a given number of samples, a good choice of q.f. can result in an accurate vector space model. Moreover, the quadrature coefficients can affect the stability of the model and decrease the condition number.

The general form of a q.f., when the derivatives are not available, is given by

$$\int_a^b f(x)dx = \sum_{i=1}^{n} c_i f(x_i) + Rf \qquad (3.8)$$

where c_i and x_i are coefficients and nodes of the q.f., respectively. *The term Rf is a functional which for any given function f(·)* equals the difference between the exact value of the integral and its approximation. For a given q.f., Rf depends on the integrand and may vanish for some specific class of functions. Therefore, the objective is to minimize the upper bound of Rf as well as to enlarge the class of functions which result in zero value for Rf. If the nodes of the q.f. are pre-assigned, the only available parameters to be treated are the coefficients. Examples of this type are Newton-Cotes and best q.f. in the sense of Sard [31]. If the nodes are free, the best location of the nodes, in a certain sense, can be determined, and the q.f. is called optimal. Examples of the optimal type are Gauss-Legendre and optimal q.f. in the sense of Sard. Since in most cases,

particularly in image processing, the location of the nodes are pre-assigned, only fixed node q.f. are considered here. Newton-Cotes q.f. is briefly studied in this section, best q.f. in the sense of Sard in section 3.5 and the experimental results are compared in section 3.6.

The basic idea of Newton-Cotes q.f. is to interpolate the sampled data by Lagrange method and then integrate it [32]. Clearly the remainder of the integral has the property that $RF = 0$ if $f \in \pi_{n-1}$, where π_{n-1} is the entire class of polynomials of degree less than or equal to n-1. This property may be used to determine the coefficients $c_1, \ldots, c_n$. A linear system of equations can be obtained by assuming $Rf = 0$ when $f(x) = 1, x, \ldots, x^{n-1}$ in equation (3.8). The coefficients are the solution of this linear system of equations.

3.3 Spline Functions

Spline functions are a class of piecewise polynomial functions satisfying continuity properties only slightly less strigent than those of polynomials, and thus they are a natural generalization of polynomials [33]. Given a strictly increasing sequence of real numbers $x_1, x_2, \ldots, x_n$, a spline function $S(x)$ of degree m with the knots $x_1, x_2, \ldots, x_n$ is a function having the following two properties:

1) In each interval (x_i, x_{i+1}), $S(x)$ is given by some polynomial of degree m or less.

2) $S(x)$ and its derivatives of order $1, 2, \ldots, m-1$ are

continuous everywhere.

When m = 0, Condition 2 is not operative, and a spline function of degree zero is a step function. A spline function of degree one is a polygon.

In general, the polynomials representing S(x) in adjacent intervals (x_i, x_{i+1}) and (x_{i+1}, x_{i+2}) are different, although this is not a requirement. S(x) might be represented by a single polynomial on the entire real line. In other words, all the polynomials of degree m or less are included in the class of spline functions satisfying the above properties. Spline functions can equally be defined as the following:

1) For $m > 0$, a spline function of degree m is a function in the class of m-1 times differentiable functions (C^{m-1}) whose m^{th} derivative is a step function.

2) A spline function of degree m is any m^{th} order indefinite integral of a step function.

Polynomials, because of their simple mathematical properties, have been widely used for interpolation and approximation. However, a polynomial fitted to a fairly large number of data points has numerous and severe undulations. There is now considerable evidence that spline functions in many situations are more adaptable approximating functions than polynomials with a comparable number of parameters. Moreover, they have been shown to be the solution of some optimization

problems [34], [35], [36].

The basis for the class of spline functions of degree m having knots $x_1, x_2, \ldots, x_m$ is given by

$$\{1, x, \ldots, x^m, (x-x_1)_+^m, (x-x_2)_+^m, \ldots, (x-x_n)_+^m\} \quad (3.9)$$

where

$$(x-x_i)_+^m = \begin{cases} 0 & \text{if } x \le x_i \\ (x-x_i)^m & \text{if } x > x_i \end{cases} . \quad (3.10)$$

Using this basis for interpolation and approximation turns out to be unstable in practice, since the matrix of the system is very badly conditioned unless m and n are both small [37]. The numerical instability is related to the mathematical properties of the truncated power functions. This difficulty can be overcome by adopting another basis for the class of spline functions. The most desirable basis consists of splines with finite support containing a minimum number of knots. This basis, called B-splines, has minimal support for a given degree and has been studied by Curry and Schoenberg [38]. A B-spline $M_i(x)$ of degree m with knots $x_i, x_{i+1}, \ldots, x_{i+m+1}$ is given by

$$M_i(x) = (m+1) \sum_{\ell=i}^{i+m+1} \frac{(x-x_\ell)_+^m}{\omega(x_\ell)} \quad (3.11)$$

where

$$w(x_\ell) = (-1)^{m+1} \prod_{\substack{j=i \\ j \neq \ell}}^{i+m+1} (x_\ell - x_j). \qquad (3.12)$$

If the knots are uniformly spaced, B-splines have the following properties:

1) The functions are shift-invariant, i.e.

$$M_i(x) = M_{i-k}(x - x_k)$$

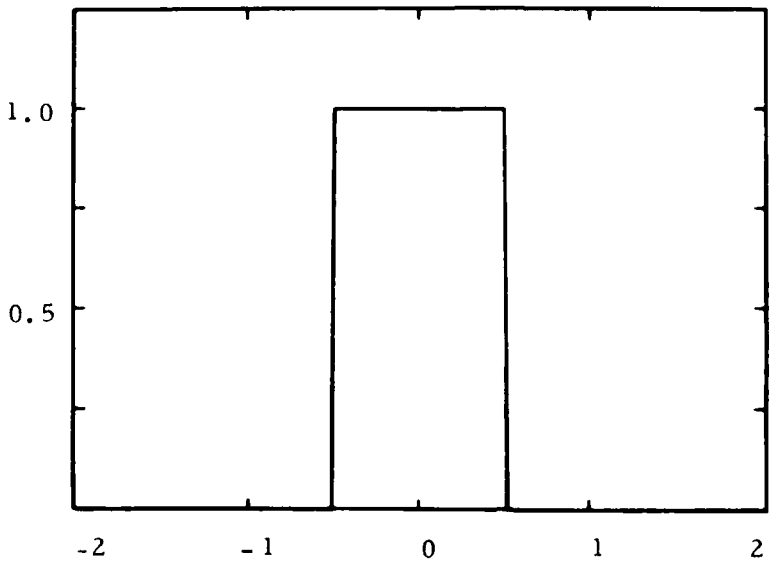
2) The functions are strictly positive.

3) The convolution of two B-splines of degrees m and n yields another B-spline of degree m+n+1.

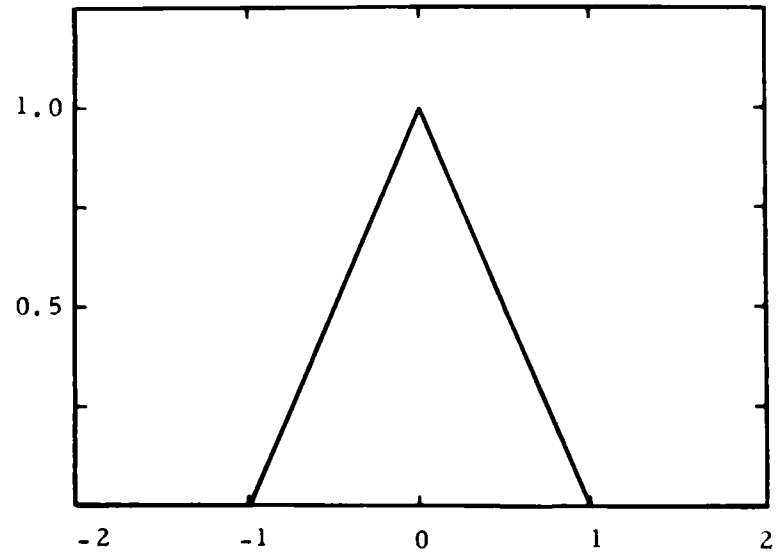
4) The functions have limited support and thus are a local basis.

These interesting properties can be fully exploited in image processing. Property 1 results in circulant matrices which can be inverted by a Fourier transform. The second property is useful because image intensity is always non-negative. Property 3 is analogous to convolution in space-invariant degradation which will be the subject of section 4.1. Finally, the fourth property produces banded matrices which are very efficient in computation.

Figures 3-1 and 3-2 are plots of B-splines of degrees 1, 2 and 3 centered at the origin of the coordinate system. A B-spline of degree zero is a rectangle function and a B-spline of degree one is a triangle

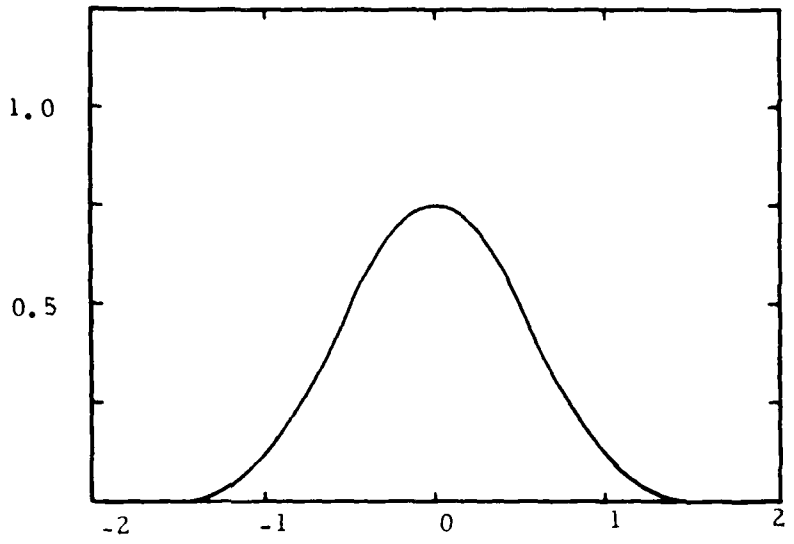


(a) B-spline of degree 0

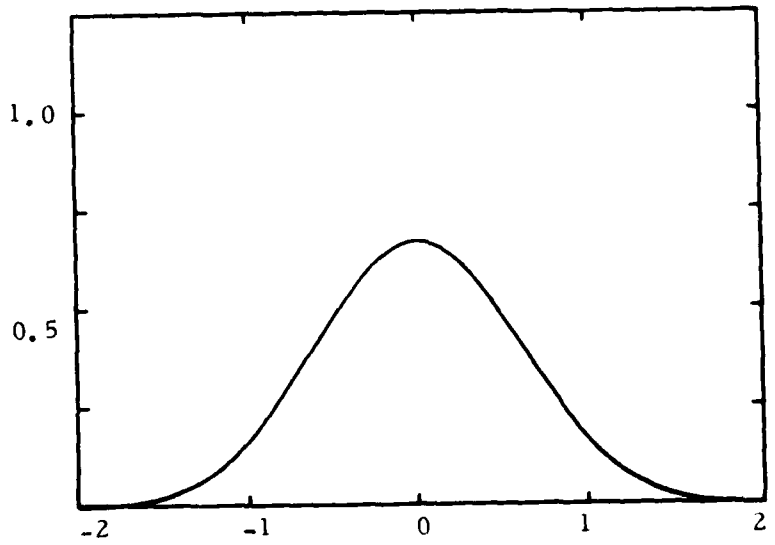


(b) B-spline of degree 1

Figure 3-1. B-splines of degrees 0 and 1.



(a) B-spline of degree 2



(b) B-spline of degree 3 (cubic B-spline)

Figure 3-2. B-splines of degree 2 and 3.

function. Using Eqs. (3.11) and (3.12), B-splines of degrees 2 and 3 can be expressed as

$$B_2(x) = 3\left[\frac{(x+1.5)_+^2}{6} - \frac{(x+.5)_+^2}{2} + \frac{(x-.5)_+^2}{2} - \frac{(x-1.5)_+^2}{6}\right] \quad (3.13)$$

$$B_3(x) = 4\left[\frac{(x+2)_+^3}{24} - \frac{(x+1)_+^3}{6} + \frac{(x)_+^3}{4} - \frac{(x-1)_+^3}{6} + \frac{(x-2)_+^3}{24}\right] \quad (3.14)$$

Spline functions of degree three, called cubic splines, in many situations have more desirable properties than other splines, and therefore they are widely used for approximation and interpolation.

3.4 Error Analysis

Interpolation and approximation by spline functions is generally not without error. For a given function, the error depends on the function, the degree of spline, and the number of placement of the knots. An analysis of the error is helpful in choosing the proper spline function. The following is a brief error analysis.

Let $PC^{m,\rho}(a,b)$ be the set of all real-valued functions $f(x)$ such that:

1) $f(x)$ is $m-1$ times continuously differentiable on the open interval (a,b).

2) There exist a sequence of knots $a = x_0 < x_1 < x_2 < \ldots < x_n < x_{n+1} = b$ such that on each open interval (x_i, x_{i+1}), $0 \le i \le n$, f is m times continuously differentiable.

3) The L_ρ-norm of m^{th} derivative is finite, i.e.,

$$\|D^m f\|_\rho \triangleq \left(\sum_{i=0}^{n} \int_{x_i}^{x_{i+1}} |D^m f(x)|^\rho dx\right)^{1/\rho} < \infty . \quad (3.15)$$

When $\rho = \infty$, the third requirement becomes

$$\|D^m f\|_\infty = \max_{0\le i\le n} \sup_{x\in(x_i, x_{i+1})} |D^m f(x)| < \infty . \quad (3.16)$$

With the above definitions, Shultz [39] has derived error bounds for different functions interpolated by cubic splines. Assuming $f(x)$ to be the original function and $S(x)$ to corresponding cubic spline function, the error bound is given as follows.

If $f\in PC^{2,2}(a,b)$, then

$$\|f-S\|_2 \le 2\pi^{-2} h^2 \|D^2 f\|_2 . \quad (3.17a)$$

If $f\in PC^{2,\infty}(a,b)$ then

$$\|f-S\|_\infty \le \frac{2}{3} h^2 \|D^2 f\|_\infty . \quad (3.17b)$$

If $f\in PC^{4,2}(a,b)$, then

$$\|f-S\|_2 \le 4\pi^{-4} h^4 \|D^4 f\|_2 . \quad (3.17c)$$

If $f\in PC^{4,\infty}(a,b)$, then

$$\|f-S\|_\infty \le \frac{5}{384} h^4 \|D^4 f\|_\infty . \quad (3.17d)$$

Inequalities (3.17a-d) indicate that the error bound is a monotonic increasing function of sampling interval h and the norm of the m-th derivative of the original function. Therefore, one way to reduce the error is to sample the function at a higher rate. When

the function f belongs to more than one of the above categories, the error bound is the smallest one. If f is a polynomial of degree less than or equal to 3, then

$$\|D^4 f\|_2 = \|D^4 f\|_\infty = 0 ,$$

therefore

$$\|f-S\|_2 = \|f-S\|_\infty = 0$$

and

$$f(x) = S(x)$$

which means cubic splines exactly interpolate the polynomials of degree 3. As another example, let $f(x)$ be a sine function with frequency u, then

$$D^2 f = -(2\pi u)^2 \sin 2\pi ux$$

and

$$D^4 f = (2\pi u)^4 \sin 2\pi ux .$$

Therefore

$$\|D^2 f\|_\infty = (2\pi u)^2$$

and

$$\|D^4 f\|_\infty = (2\pi u)^4 .$$

Substituting the above norms in (3.17b,d)

$$\|f-S\|_\infty \leq \frac{2}{3} (2\pi hu)^2$$

and

$$\|f-S''\|_\infty \le \frac{5}{384}(2\pi hu)^4 .$$

The error bound is the minimum of the above two limits. For a given error bound, h is proportional to the inverse of u which is similar to the sampling theory studied in section 3.1.

3.5 Monospline and Best Quadrature Formulae in the Sense of Sard

Monosplines are a class of functions defined as [31]

$$K(x) = \frac{x^m}{m!} - S_{m-1,n}(x) \tag{3.18}$$

where $S_{m-1,n}(x)$ is a spline function of degree m-1 with n pre-assigned knots $a < x_1 < x_2 < \ldots < x_n < b$ and $n > m$ given by

$$S_{m-1,n}(x) = \sum_{j=0}^{m-1} \alpha_j x^j + \sum_{i=1}^{n} \frac{c_i (x-x_i)_+^{m-1}}{(m-1)!} . \tag{3.19}$$

K(x) which consists of a polynomial of degree m and a spline of degree m-1 is called a monospline of degree m with n nodes. Using $K^{(m)}(x)$ as a kernel

$$\int_a^b f(x)K^{(m)}(x)dx = \int_a^b f(x)dx - \sum_{i=1}^{n} \int_a^b c_i f(x)\delta(x-x_i)dx$$

$$= \int_a^b f(x)dx - \sum_{i=1}^{n} c_i f(x_i) . \tag{3.20}$$

Integrating the left hand side of (3.20) m times (by parts) gives

$$\int_a^b f(x)K^{(m)}(x)dx = \sum_{j=0}^{m-1} (-1)^j f^{(j)}(x)K^{(m-1-j)}(x)\Big|_a^b$$

$$+ (-1)^m \int_a^b f^{(m)}(x)K(x)dx . \tag{3.21}$$

Assuming

$$K^{(j)}(a) = K^{(j)}(b) = 0 \quad \text{for } j = 0, 1, \ldots, m-1 \tag{3.22}$$

and substituting (3.21) in (3.20) gives

$$\int_a^b f(x)dx = \sum_{i=1}^{n} c_i f(x_i) + (-1)^m \int_a^b f^{(m)}(x)K(x)dx. \tag{3.23}$$

Therefore the remainder of the integral, or the error of q.f., is given by

$$Rf = (-1)^m \int_a^b f^{(m)}(x)K(x)dx \tag{3.24}$$

The upper bound of Rf can be expressed as

$$|Rf| = \left| \int_a^b f^{(m)}(x)K(x) \right| \leq \|K\|_2 \|f^{(m)}\|_2 \tag{3.25}$$

where $\|\cdot\|_2$ denotes L_2-norm of the function. If Rf = 0 when f is a polynomial of degree less than or equal to m 1 and K(x) has the least square deviation (minimum norm) among all kernels of the form (3.18), then the q.f. is called the best in the sense of Sard. For a given function f, the minimum norm of K generates the minimum upper bound of Rf. Assumption (3.22) which leads to Eq. (3.23) satisfies the first requirement. Schoenberg [40], [41] has shown that there exists a unique monospline

$$H(x) = \frac{x^{2m}}{(2m)!} - S_{2m-1,n}(x) \tag{3.26}$$

of degree 2m in which the kernel K(x) of Sard's best q.f., in terms of

H(x), is given by

$$K(x) = H^{(m)}(x) \ . \tag{3.27}$$

H(x) must satisfy the following conditions

$$H(x_i) = 0 \qquad i = 1, 2, \ldots, n \tag{3.28a}$$

$$H^{(m+j)}(a) = 0 \qquad j = 0, 1, \ldots, m-1 \tag{3.28b}$$

$$H^{(m+j)}(b) = 0 \qquad j = 0, 1, \ldots, m-1 \tag{3.28c}$$

The value J of minimum derivation $\|K\|_2^2$, in terms of H(x), is determined by the relation

$$J = \int_a^b [K(x)]^2 dx = (-1)^m \int_a^b H(x)dx \ . \tag{3.29}$$

Assuming a = -1, b = 1 (this assumption can always be made by the normalizing $S = \frac{x-a}{b-a}$), and applying condition (3.28b), H(x) can be written as

$$H(x) = \frac{(x+1)^{2m}}{(2m)!} - \sum_{j=0}^{m-1} \alpha_j x^j - \sum_{j=1}^{n} \frac{c_j (x-x_j)_+^{2m-1}}{(2m-1)!} \ . \tag{3.30}$$

Conditions (3.28a) and (3.28c) generate a system of m+n linear equations with m+n unknowns of the form

$$\frac{(x_i+1)^{2m}}{(2m)!} - \sum_{j=0}^{m-1} \alpha_j x_i^j - \sum_{j=1}^{n} \frac{c_j (x_i-x_j)_+^{2m-1}}{(2m-1)!} = 0 \qquad i=1, 2, \ldots, n \tag{3.31a}$$

$$\frac{2^i}{i} - \sum_{j=1}^{n} c_j (1-x_j)^{i-1} = 0 \qquad i = 1, \ldots, m \ . \tag{3.31b}$$

The coefficients c_j and α_j and the L_2-norm of K can be obtained by solving the above linear system of equations. If m = 1, which corresponds to the pulse approximation method, the system can easily be solved. In this case the equations are given by

$$\frac{(x_i+1)^2}{2} - \alpha_0 - \sum_{j=1}^{n} c_j(x_i - x_j)_+ = 0 \qquad i = 1,2,\ldots,n \qquad (3.32a)$$

$$\sum_{i=1}^{n} c_i = 2 \quad . \qquad (3.32b)$$

The solution to the above equation is obtained as

$$\alpha_0 = \frac{(1+x_1)^2}{2} \qquad (3.33a)$$

$$c_1 = \frac{2+x_1+x_2}{2} \qquad (3.33b)$$

$$c_j = \frac{x_{j+1}-x_{j-1}}{2} \qquad j = 2,3,\ldots,n-1 \qquad (3.33c)$$

$$c_n = \frac{2-x_{n-1}-x_n}{2} \quad . \qquad (3.33d)$$

When the sample points are equidistant, the location of nodes and coefficients are derived as follows

$$x_j = -1 + \frac{2j-1}{n} \qquad j = 1,2,\ldots,n \qquad (3.34a)$$

$$c_j = \frac{2}{n} \qquad j = 1,2,\ldots,n \qquad (3.34b)$$

$$\alpha_0 = \frac{1}{2n^2} \quad . \qquad (3.34c)$$

Substituting (3.34a-c) in (3.30) produces

$$H(x) = \frac{(x+1)^2}{2} - \frac{1}{2n^2} - \frac{2}{n}\sum_{j=1}^{n}(x+x_j)_+ . \qquad (3.35)$$

The value J of minimum deviation can be obtained as

$$\begin{aligned} J &= \int_{-1}^{1}(K(x))^2 dx = (1-)^m \int_{-1}^{1} H(x)dx \\ &= -\int_{-1}^{1}\frac{(x+1)^2}{2}\,dx + \frac{1}{2n^2}\int_{-1}^{1}dx + \frac{2}{n}\sum_{j=1}^{n}\int_{-1}^{1}(x-x_j)_+ dx \\ &= -\frac{8}{6} + \frac{1}{n^2} + \frac{1}{n}\sum_{j=1}^{n}\left(2 + \frac{1}{n} - \frac{2j}{n}\right)^2 \\ &= \frac{2}{3n^2} \quad . \end{aligned}$$

Substituting the norm of K(x) in (3.25), the upper bound of the error is

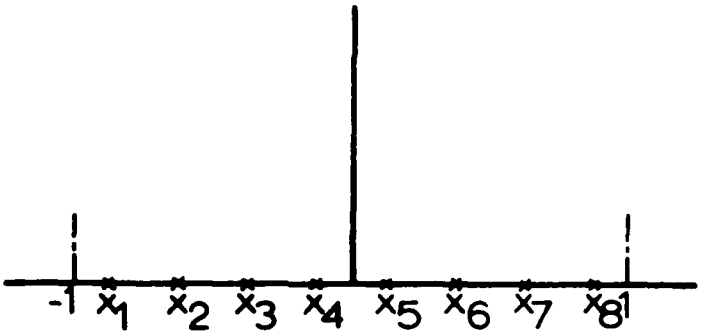
$$Rf \le \frac{1}{n}\sqrt{\frac{2}{3}}\,\|f'\| \qquad (3.36)$$

Therefore, as was expected, the upper bound of error is inversely proportional to the number of sample points and approaches zero as n increases. Moreover, Rf = 0 for constant functions f(x) = c regardless of the number of sample points.

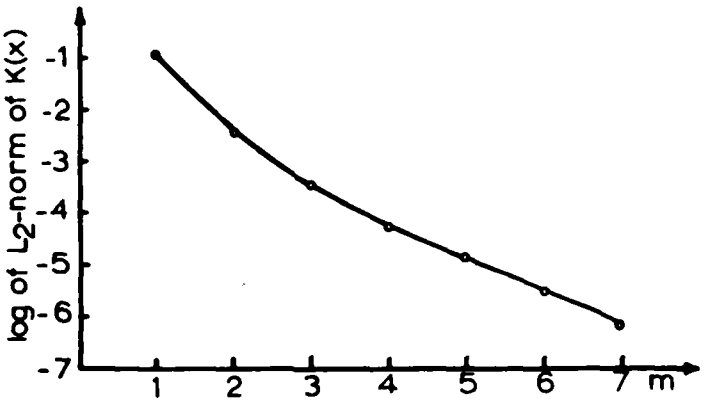
3.6 Experimental Results

To show the improvements that can be made by using monosplines, this section is devoted to applying Sard's best q.f. to a variety of functions and comparing the results with pulse approximation method and Newton-Cotes q.f.

In applying Sard's best q.f., one is faced with the task of selecting the parameter m. For a given number of sample points, the deviation of K(x) decreases as m increases. Figure 3-3 is an illustration of this property for 8 uniformly spaced nodes. Therefore, one may assume m to be the highest order where $f^{(m)}$ is continuous. On the other hand, as m increases, different problems will arise. First, the system of equations (3.31a, b) tends to become unstable for large m. Second, in some cases, the norm of the m-th derivative of the integrand increases rapidly as m grows, and a smaller choice of m would result in a smaller error. To study the effects of different values of m on the error and also to compare Sard's method to the Newton-Cotes and pulse approximation methods, several experiments have been performed. Figure 3-4a demonstrates the error as a function of frequency for a sine function. Case m =1 coincides with the pulse approximation method, and m = 8 is equivalent to the Newton-Cotes q.f. Figure 3-4b is a plot of the theoretical error bound for a sine function. In Figure 3-5a the integrand is a polynomial of degree 8. Variable j is a measure of how fast the polynomial oscillates in the interval of integration; j is almost equal to the number of roots in the interval (-1, 1) minus one. In other words, the larger j becomes, the harder it is to approximate the function since it is subject to more fluctuation. This roughly corresponds to the frequency in Fig. 3-4. Figure 3-5b shows the

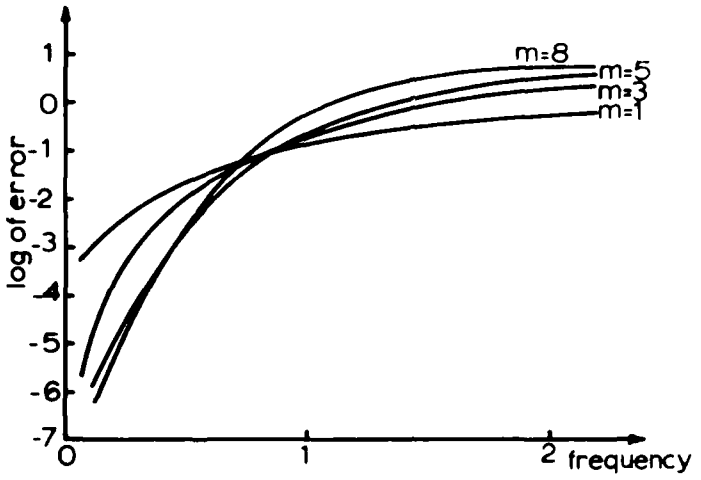


(a) Uniform spacing of 8 nodes

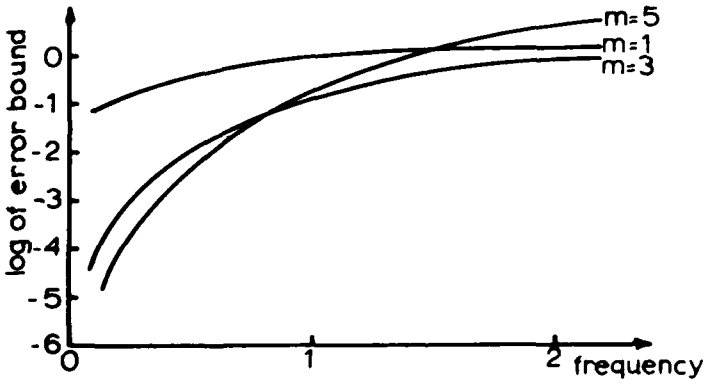


(b) L_2-norm of K(x) vs. degree of monospline

Figure 3-3. Deviation of the kernel for 8 uniform nodes.

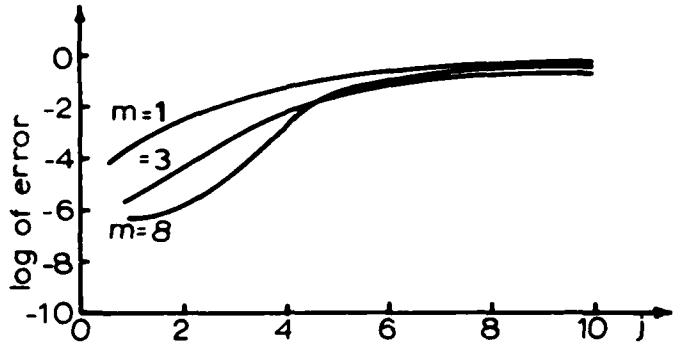


(a) Error

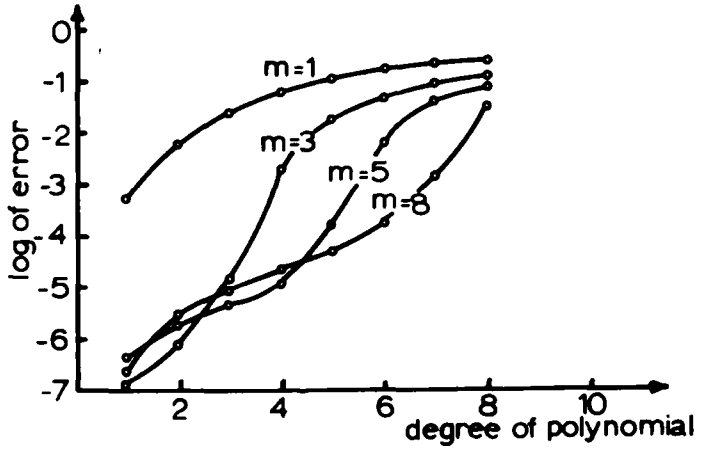


(b) Error bound

Figure 3-4. Quadrature error and error bound for sine functions.



(a) Error for polynomial of degree 8



(b) Error for polynomials of various degrees

Figure 3-5. Quadrature error for polynomials.

error for polynomials of different degrees. In this plot all the roots are between -1 and 1.

Both theory and experience indicate that the choice of m greatly depends on the frequency content of the integrand f. For the class of rapidly varying functions, a smaller m is advised, but for the class of slowly varying functions, large values of m give better results. Since it is assumed that f is sampled faster or equal to the Nyquist rate, the curves of Fig. 3-4 are not studied for frequencies above two cycles. Figures 3-4 and 3-5 show that the curves cross each other and that a tradeoff exists between the frequency content of the integrand and degree of monospline. Considering this fact and taking into account the set of examples, the cubic monospline produces less error overall and thus the optimal value for m is three. Of course, for other cases where the function f is highly oversampled, large values of m may be recommended, while on the other hand, when the function is sampled far below the Nyquist rate, the pulse approximation is preferred to the other techniques.

In section 4.4, Sard's best q.f. has been used in the simulation of images degraded by astigmatism and curvature of the field. This q.f. results in a more accurate model with the reduction of simulation artifacts. Moreover, this technique has decreased the condition number of the blur matrix and consequently produced a more stable model [24], [42].

Chapter 4

RESTORATION OF NOISELESS IMAGES

The restoration of noise-free images is presented in this chapter. The convolutional property of B-splines is used for the restoration of space-invariant degradations. It is shown that representing the object and point-spread function by B-splines leads to a more accurate reconstruction of the original object than the conventional method.

The singularity of most imaging systems due to the irreversible loss of original object information is a major problem in image restoration. A minimum norm principle leading to pseudo-inversion is used to overcome this difficulty. This technique is applicable to space-variant degradations, underdetermined models and overdetermined models. The space-variant point-spread functions that describe imaging in the presence of astigmatism and curvature of field are derived and coordinate transformations are applied to reduce the dimensionality. The singular-value-decomposition is used for solution of the simplified equations.

4.1 Application of B-splines to Space-Invariant Degradations

As discussed in Chapter 2, the deterministic part of a degraded image in a space-invariant imaging system is described by a convolution integral. Using B-splines as a basis in uniform

sampling, the object f(x) and point-spread function h(x) can be represented in the forms

$$f(x) = \sum_i f_i B_m(x-x_i) \qquad (4.1)$$

$$h(x) = \sum_j h_j B_n(x-x_j) \qquad (4.2)$$

where $B_m(x)$ and $B_n(x)$ are B-splines of degrees m and n centered at the origin, and f_i and h_j are interpolation coefficients. Substituting (4.1,2) in the convolution integral, the image is

$$g(x) = \int_{-\infty}^{\infty} h(x-\xi)f(\xi)d\xi$$

$$= \sum_i \sum_j f_i h_j B_m(x-x_i) * B_n(x-x_j) . \qquad (4.3)$$

Exploiting the convolutional property of B-splines

$$B_m(x-x_i) * B_n(x-x_j) = B_{m+n+1}(x-x_i-x_j) \qquad (4.4)$$

and representing g(x) by a B-spline of degree m+n+1, Eq. (4.3) can be written in the form

$$\sum_k g_k B_{m+n+1}(x-k\Delta x) = \sum_i \sum_j f_i h_j B_{m+n+1}(x-(i+j)\Delta x) \qquad (4.5)$$

where Δx is the sampling interval. Equations (4.3) and (4.5) show that the B-spline, which is interpolating the deterministic part of the blurred image, must be of higher degree than the B-splines

interpolating the object and point-spread function. In other words, since the blurred image is always smoother than the object, a higher degree spline can follow the image function better than one approximating the object function. This can be explained in the Fourier domain by observing that the Fourier transform of a m^{th} degree B-spline is a sine function to the power m+1. As m increases the amplitude of higher frequencies decreases. Since a blurred image has less higher frequency content than the object, a higher order B-spline can represent the image better than the one representing the object.

Using vector space notation, Eq. (4.5) may be written as

$$\underline{g} = \underline{H}\,\underline{f} \tag{4.6}$$

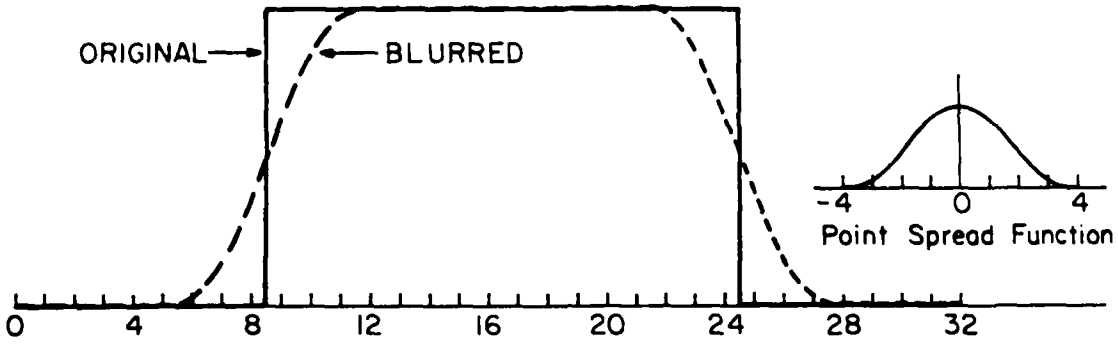
where $\underline{g}$ and $\underline{f}$ are vectors consisting of coefficients g_k and f_i, and $\underline{H}$ is a circulant matrix with elements h_j. If the point-spread-function is of finite width, the matrix $\underline{H}$ is banded.

As an experiment to compare spline functions with the pulse approximation method, a rectangular object is blurred analytically by a 4^{th} degree polynomial of the form

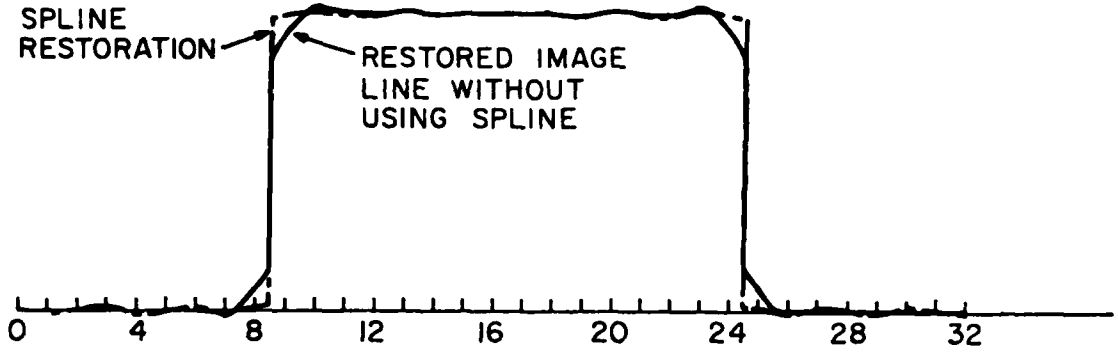
$$h(x) = \frac{15}{56}\left(1 - \left(\frac{x}{3.5}\right)^2\right) \qquad -3.5 \le x \le 3.5 \tag{4.7}$$

$$= 0 \quad , \qquad \text{elsewhere}$$

and this is plotted in Fig. 4-1. The object is a rectangular function,



(a) Original and blurred image line



(b) Restored image line with and without using spline

Figure 4-1. Restoration of space-invariant degradation using spline functions.

therefore it is interpolated by B-splines of degree zero. The second derivative of h at points $x = -3.5$ and $x = 3.5$ is a step function and it is interpolated by B-splines of degree two. Since the convolution of a zero degree and a second degree B-spline is a cubic spline, the image is interpolated by cubic B-splines. Figure 4-1b, the restored image with and without splines, shows that the spline restores the edges much sharper and generates less undulations than the common pulse approximation method. Using different degrees of B-splines for object, image and point-spread functions depending on their characteristic has led to a more accurate model and thus a better quality restoration. Spline restoration can also be applied to a two-dimensional blur with very good results [43].

4.2 Restoration of Space-Variant Degradations by the Minimum Norm Principle

In the previous chapter, a noiseless blurred image was modeled by the expression

$$\underline{g} = \underline{H}\,\underline{f} \tag{4.8}$$

where $\underline{H}$ represents the blurred matrix. If $\underline{H}$ is square, non-singular and well-conditioned, the restored image $\hat{\underline{f}}$ can be obtained by

$$\hat{\underline{f}} = \underline{H}^{-1}\underline{g} \tag{4.9}$$

where $\underline{H}^{-1}$ denotes the inverse of $\underline{H}$. In most practical cases, $\underline{H}$ is either singular or ill-conditioned due to its large size and due to the fact that most imaging systems irreversibly remove certain aspects

of the original object. In underdetermined or overdetermined models, which will be defined later, $\underline{H}$ is not a square matrix. Thus even in the absence of noise an estimate of $\underline{f}$ cannot be obtained by (4.9). This suggests the definition of a reasonable fidelity criteria which leads to a unique solution for $\hat{\underline{f}}$. The minimum norm criterion is defined as the following

$$\text{minimize } \|\underline{f}\|^2 \tag{4.10}$$

among all $\underline{f} \varepsilon R^n$ which minimizes

$$\|\underline{g}-\underline{H}\,\underline{f}\|^2 \tag{4.11}$$

where $\|\cdot\|$ denotes L_2-norm of the vector. Albert [44] has shown that there exists a unique solution for the above minimization problem. The solution to (4.10) and (4.11) may be obtained by the standard methods of the calculus of variations. Using Lagrangian parameter δ^2, the functional

$$W(\underline{f}) = \|\underline{g}-\underline{H}\,\underline{f}\|^2 + \delta^2\|\underline{f}\|$$

must be minimized. Taking a derivative with respect to $\underline{f}$,

$$\frac{\partial W}{\partial \underline{f}} = -2\underline{H}^t(\underline{g}-\underline{H}\,\underline{f}) + 2\delta^2\underline{f} = 0$$

the optimal estimate for $\underline{f}$ is

$$\hat{\underline{f}} = \lim_{\delta \to 0} (\underline{H}^t\underline{H} + \delta^2 I)^{-1}\underline{H}^t\underline{g} \tag{4.12}$$

where I represents the identity matrix whose dimensionality is

understood from the context. For example, if $\underline{H}$ is m x n, I is an n x n identity matrix. It is shown [44] that for any m x n matrix $\underline{H}$

$$\underline{H}^{+} = \lim_{\delta \to 0} (\underline{H}^{t}\underline{H} + \delta^{2}I)^{-1}\underline{H}^{t} \tag{4.13}$$

always exists and $\underline{H}^{+}$ is called the pseudoinverse of H. For any m dimensional vector $\underline{g}$,

$$\hat{\underline{f}} = \underline{H}^{+}\underline{g} \tag{4.14}$$

is the vector of minimum norm among those which satisfy (4.11). The minimum norm $\hat{\underline{f}}$ is an element of $\mathcal{R}(\underline{H}^{t})$, the range of $\underline{H}^{t}$, and satisfies the relation

$$\underline{H}\,\hat{\underline{f}} = \hat{\underline{g}}$$

where $\hat{\underline{g}}$ is the projection of $\underline{g}$ on $\mathcal{R}(\underline{H})$. Since

$$(\underline{H}^{t}\underline{H}\,\underline{H}^{t} + \delta^{2}\underline{H}^{t}) = \underline{H}^{t}(\underline{H}\,\underline{H}^{t} + \delta^{2}I) = (\underline{H}^{t}\underline{H} + \delta^{2}I)\underline{H}^{t}$$

and since $(\underline{H}\,\underline{H}^{t} + \delta^{2}I)$ and $(\underline{H}^{t}\underline{H} + \delta^{2}I)$ have inverses when $\delta^{2} > 0$, it is clear that

$$(\underline{H}^{t}\underline{H} + \delta^{2}I)^{-1}\underline{H}^{t} = \underline{H}^{t}(\underline{H}\,\underline{H}^{t} + \delta^{2}I)^{-1}$$

and $\underline{H}^{+}$ can also be expressed as

$$\underline{H}^{+} = \lim_{\delta \to 0} \underline{H}^{t}(\underline{H}\,\underline{H}^{t} + \delta^{2}I)^{-1} \tag{4.15}$$

Equivalently, the pseudoinverse of a m x n matrix $\underline{H}$ is defined as an n x m matrix $\underline{X}$ satisfying the following four properties:

$$1)\quad \underline{H}\,\underline{X}\,\underline{H} = \underline{H}, \tag{4.16a}$$

$$2)\quad \underline{X}\,\underline{H}\,\underline{X} = \underline{X}, \tag{4.16b}$$

$$3)\quad (\underline{H}\,\underline{X})^t = \underline{H}\,\underline{X}, \tag{4.16c}$$

$$4)\quad (\underline{X}\,\underline{H})^t = \underline{X}\,\underline{H}. \tag{4.16d}$$

The above properties are necessary and sufficient conditions for $\underline{X} = \underline{H}^+$ given by (4.13) or (4.15).

When object and image are represented by other basis functions, such as B-splines, in a continuous-continuous model, a similar minimization criterion may be applied. Let

$$f(\xi) = \sum_{i=1}^{M} f_i B_m(\xi - \xi_i) \tag{4.17a}$$

$$g(x) = \sum_{i=1}^{N} g_i B_n(x - x_i) \tag{4.17b}$$

where B_m and B_n are B-splines of degrees m and n. Here g(x) is related to f(x) by the superposition integral given by (3.5). Defining the following objective function

$$W(\underline{f}) = \int_{-\infty}^{\infty} \left(g(x) - \int_{-\infty}^{\infty} h(x, \xi) f(\xi) d\xi\right)^2 dx + \delta^2 \int (f(\xi))^2 d\xi, \tag{4.18}$$

substituting (4.17a,b) into (4.18) and taking derivatives with respect to $\underline{f}$, the optimal estimate for $\underline{f}$ is

$$\hat{\underline{f}} = (\underline{P} + \delta^2 \underline{B})^{-1} \underline{Q}\,\underline{g} \tag{4.19}$$

where the vectors $\hat{\underline{f}}$, $\underline{g}$, and the matrices $\underline{P}$, $\underline{B}$ and $\underline{Q}$ are defined by

$$\hat{\underline{f}} = [\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_m]^t$$

$$\underline{g} = [g_1, g_2, \ldots, g_N]^t$$

$$\underline{B}_m(\xi) = [B_m(\xi-\xi_1), B_m(\xi-\xi_2), \ldots, B_m(\xi-\xi_M)]^t$$

$$\underline{B}_n(x) = [B_n(x-x_1), B_n(x-x_2), \ldots, B_n(x-x_N)]^t$$

$$\underline{p}(x) = \int_{-\infty}^{\infty} h(x, \xi)\underline{B}_m(\xi)d\xi$$

$$\underline{P} = \int_{-\infty}^{\infty} \underline{p}(x)\underline{p}^t(x)dx$$

$$\underline{B} = \int_{-\infty}^{\infty} \underline{B}_m(\xi)\underline{B}_m^t(\xi)d\xi$$

$$\underline{Q} = \int_{-\infty}^{\infty} \underline{p}(x)\underline{B}_n^t(x)dx$$

The matrix $\underline{P}$ is symmetric and non-negative definite. Assuming $\underline{q}$ to be an arbitrary vector of dimension M, then

$$\underline{q}^t\underline{P}\underline{q} = \int_{-\infty}^{\infty} \underline{q}^t\underline{p}(x)\underline{p}^t(x)\underline{q}dx = \int_{-\infty}^{\infty} (\underline{q}^t\underline{p}(x))^2dx \geq 0.$$

The matrix B is symmetric, banded and positive definite matrix consisting of the values of a B-spline of degree 2n+1 at its knots. Therefore, $(\underline{P}+\delta^2\underline{B})$ is positive definite and invertible.

If the functions $\underline{f}$, $\underline{g}$ and $\underline{h}$ are represented by B-splines of degree zero, then $\underline{P} = \underline{H}^t\underline{H}$, $\underline{Q} = \underline{H}^t$ and $\underline{B} = I$, and Eq. (4.19) simplifies to (4.12). A similar formula is derived for the continuous-discrete model when the objective function is defined by the minimization of

the second derivative of f [45].

4.3 Pseudoinversion and Singular-Value-Decomposition

The specific structure and properties of a matrix are quite useful in determining its pseudoinverse. For a nonsingular square matrix, since $\hat{\underline{f}} = \underline{H}^{-1}\underline{g}$ is the only vector minimizing (4.11), the pseudoinverse is the same as the inverse. If matrix $\underline{H}$ is diagonal:

$$\underline{H} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \tag{4.20}$$

then

$$\underline{H}^{+} = \text{diag}(\lambda_1^{+}, \lambda_2^{+}, \ldots, \lambda_n^{+}) \tag{4.21}$$

where

$$\lambda_i^{+} = \begin{cases} \lambda_i^{-1} & \text{if } \lambda_i \neq 0 \\ 0 & \text{if } \lambda_i = 0 \end{cases} \qquad i = 1, \ldots, n \; . \tag{4.22}$$

This result agrees with the result from a least squares viewpoint. For $\underline{H}$ given by (4.20), the value of

$$\|\underline{g} - \underline{H}\,\underline{f}\|^2 = \sum_{i=1}^{n} (g_i - \lambda_i f_i)^2$$

is minimum when

$$\hat{f}_i = \begin{cases} \lambda_i^{-1} g_i & \text{if } \lambda_i \neq 0 \\ \text{arbitrary} & \text{if } \lambda_i = 0 \end{cases} \qquad i = 1, \ldots, n$$

Among all vectors $\hat{\underline{f}}$ satisfying (4.1), the one with minimum norm is

$$\hat{f}_i = 0 \quad \text{if } \lambda_i = 0.$$

Thus, the minimum norm solution for a diagonal matrix is $\hat{\underline{f}} = \underline{H}^{+}\underline{g}$

where $\underline{H}^+$ is defined by (4.21).

Equation (4.22) shows the radically discontinuous nature of pseudoinversion. Two matrices may be very close to each other element by element, but their pseudoinverses differ greatly. For example, the diagonal matrices

$$\underline{\Lambda}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \underline{\Lambda}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 10^{-5} \end{bmatrix}$$

are close to each other, but

$$\underline{\Lambda}_1^+ = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \underline{\Lambda}_2^+ = \begin{bmatrix} 1 & 0 \\ 0 & 10^{5} \end{bmatrix}$$

differ greatly. The reason is that (4.22) exhibits an infinite discontinuity at $\lambda = 0$. This characteristic induces serious computational difficulties, particularly due to computer precision and round-off error where a small number might be actually zero or vice versa. This will be discussed more in the computation of $\underline{H}^+$ for a general m x n matrix.

The pseudoinverse of a symmetric matrix can be derived by using the diagonalization theorem. A symmetric matrix $\underline{H}$ can be written as

$$\underline{H} = \underline{E}\,\underline{\Lambda}\,\underline{E}^t \tag{4.23}$$

where $\underline{E}$ is an orthogonal matrix and $\underline{\Lambda}$ is diagonal. Substituting (4.23) in (4.12) gives

$$\underline{H}^{+} = \lim_{\delta \to 0} \underline{E}(\underline{\Lambda}^2 + \delta^2 I)^{-1} \underline{\Lambda} \underline{E}^t$$

$$= \underline{E}[\lim_{\delta \to 0} (\underline{\Lambda}^2 + \delta^2 I)^{-1} \underline{\Lambda}] \underline{E}^t$$

$$= \underline{E}\, \underline{\Lambda}^{+}\, \underline{E}^t \qquad (4.24)$$

where $\underline{\Lambda}^{+}$ is defined by (4.21) and (4.22). Thus, the pseudoinverse for a symmetric matrix is obtained by pseudoinverting the diagonal matrix of its eigenvalues. Equation (4.24) can equally be expressed as

$$\underline{H}^{+} = \sum_{i=1}^{n} \lambda_i^{+} \underline{e}_i \underline{e}_i^t \qquad (4.25)$$

where $\underline{e}_i$ is the eigenvector of $\underline{H}$ associated with the eigenvalue λ_i.

If $\underline{H}$ is a rectangular matrix of full row rank, i.e., the rows of $\underline{H}$ are linearly independent, then $\underline{H}\,\underline{H}^t$ is invertible and eq. (4.15) simplifies to

$$\underline{H}^{+} = \underline{H}^t (\underline{H}\,\underline{H}^t)^{-1} \qquad (4.26)$$

Although Eq. (4.25) presents a straightforward method for computing $\underline{H}^{+}$, the problem of inverting the m x m matrix $(\underline{H}\,\underline{H}^t)$ remains. This can cause difficulties for even moderate size images. In this situation, the observation can be partitioned into smaller segments which are used for estimation of the corresponding object sections. Moreover, since the number of linear equations is less than the number of unknowns in (4.8), the estimated object $\hat{\underline{f}}$ is not necessarily equal to the original object $\underline{f}$. In other words, a full recovery of the

vector $\underline{f}$ is not guaranteed. Nevertheless, the estimated vector $\hat{\underline{f}}$ satisfies the relation

$$\underline{H}\,\hat{\underline{f}} = \underline{g} = \underline{H}\,\underline{f}$$

because

$$\underline{H}\,\hat{\underline{f}} = \underline{H}\,\underline{H}^{+}\underline{g} = \underline{H}\,\underline{H}^{t}(\underline{H}\,\underline{H}^{t})^{-1}\underline{g} = \underline{g}\,.$$

If $\underline{H}$ is a rectangular matrix of full column rank, i.e., the columns of $\underline{H}$ are linearly independent, then $\underline{H}^{t}\underline{H}$ is invertible and Eq. (4.15) simplifies to

$$\underline{H}^{+} = (\underline{H}^{t}\underline{H})^{-1}\underline{H} \tag{4.27}$$

In this case, the number of linear equations, or in other words, the number of observations, is more than the number of unknowns. If the system of equations is consistent, i.e., $\underline{g}$ is a linear combination of column vectors of $\underline{H}$, which is always true in noiseless models, the estimated object $\hat{\underline{f}}$ is the same as the original object f because

$$\hat{\underline{f}} = \underline{H}^{+}\underline{g} = (\underline{H}^{t}\underline{H})^{-1}\underline{H}^{t}\underline{g} = (\underline{H}^{t}\underline{H})^{-1}\underline{H}^{t}\underline{H}\underline{f} = \underline{f}\,.$$

Thus the object can be recovered from the image without any error. A computationally efficient method utilizing the Fourier properties of circulant matrices has been introduced for the pseudoinversion of full column matrices in space-invariant degradation [13].

When $\underline{H}$ is a general m x n matrix with no particular structure, singular-value-decomposition (SVD) techniques can be used for

computation of the pseudoinverse. Let $\underline{H}$ be an m x n matrix and let $\underline{\Lambda}$ be the r x r diagonal matrix consisting of the square roots of the nonzero eigenvalues of $\underline{H}\,\underline{H}^t$. Then there exists an m x r matrix $\underline{U}$ and an r x n matrix $\underline{V}$ such that the following conditions hold [44], [46]

$$\underline{H} = \underline{U}\,\underline{\Lambda}\,\underline{V}^t \tag{4.28a}$$

$$\underline{U}^t\underline{U} = \underline{V}\,\underline{V}^t = I \tag{4.28b}$$

The columns of $\underline{U}$ are orthonormal eigenvectors of $\underline{H}\,\underline{H}^t$ and the rows of $\underline{V}$ are the orthonormal eigenvectors of $\underline{H}^t\underline{H}$. The decomposition (4.28) is called the singular-value-decomposition. Equation (4.28a) can be represented as

$$\underline{H} = \sum_{i=1}^{r} \lambda_i \underline{u}_i \underline{v}_i^t \tag{4.29}$$

which leads to pseudoinverse of H in the following form

$$\underline{H}^+ = \sum_{i=1}^{r} \lambda_i^{-1} v_i u_i^t = \underline{V}\,\underline{\Lambda}^+\,\underline{U}^t . \tag{4.30}$$

The SVD algorithm developed by Golub and Reinsch [46] computes λ_i, u_i and v_i, i = 1,...,n, in a numerically stable way without explicitly forming $\underline{H}\,\underline{H}^t$ or $\underline{H}^t\underline{H}$. It uses a Housholder transformation to reduce $\underline{H}$ to a bidiagonal form, and then the QR algorithm to find the singular values of the bidiagonal matrix.

In practical cases, a judicious choice of eigenvalue cutoff λ_r must be made for nonzero eigenvalues. If the λ_i's, ordered in

decreasing value, show a sudden decrease in value as a function of the index i, then the threshold may be located at that point. The decrease in value can be by a factor as small as the machine precision. If such a sudden decrease does not exist, a threshold ε which is dependent on machine precision must be selected and the eigenvalues smaller than ε are declared zero. The value r determines the rank of $\underline{H}$ and small eigenvalues $\lambda_{r+1}, \ldots, \lambda_n$ are assumed to be roundoff error. Equation (4.29) expands $\underline{H}$ in terms of system eigenvectors; thus the λ_i's are the effective spectral components.

General outer-product expansions of $\underline{H}$ are given by

$$\underline{H} = \sum_i \sum_j a_{ij} \underline{u}_i \underline{v}_j^t \tag{4.31}$$

where u_i and v_j can be the discrete Fourier basis vectors, Walsh-Hadamard, Haar, Slant, or other orthonormal bases. With a space-invariant degradation, $\underline{H}$ is a circulant matrix that can be diagonalized by discrete Fourier transforms. Thus, the SVD procedure is analogous to the discrete Fourier-inverse-filtering method that is widely used for space-invariant processing.

4.4 Restoration of Astigmatism and Curvature-of-Field

Optical images are subject to a number of blurring effects due to aberrations. Certain aberrations, such as spherical aberration, can be described by convolution integrals and can be solved in the Fourier domain. For other aberrations such as coma, astigmatism

or curvature-of-field, the blurring is space-variant. When the effects of astigmatism and curvature-of-field predominate, the geometrical-optics aberration functions

$$x-\xi = (2C+D)\xi^2 r \cos\theta \tag{4.32a}$$

$$y-\eta = D\xi^2 r \sin\theta \tag{4.32b}$$

describe the displacement of an image point from its ideal (Gaussian) intercept in the image plane. Here r and θ are ray intercepts in the exit pupil of the optical system, and C and D are constant coefficients describing the degree of astigmatism and curvature-of-field, respectively [22], [48]. Using a technique described in [22] and [48], the space-variant point-spread function (SVPSF) of the system for a circular exit pupil of radius R is obtained as

$$h(x,y;\xi,\eta=0) = \begin{cases} \dfrac{1}{D(2C+D)\xi^4}, & \dfrac{y^2}{D^2R^2\xi^4} + \dfrac{(x-\xi)^2}{(2C+D)^2R^2\xi^4} \le 1 \\ 0, & \text{elsewhere} \end{cases} \tag{4.33}$$

assuming an object impulse function at $(\xi, \eta=0)$. This function is given in Fig. 4-2 for the impulses at various locations in the (ξ, η) plane. The region of nonzero response are defined by ellipses which increase in size proportional to the square of the radial distance, and the amplitude of the response decreases inversely with ξ^4. Although the system is strongly space-variant and the blurring occurs in both radial and angular directions, changes in the amplitude and shape

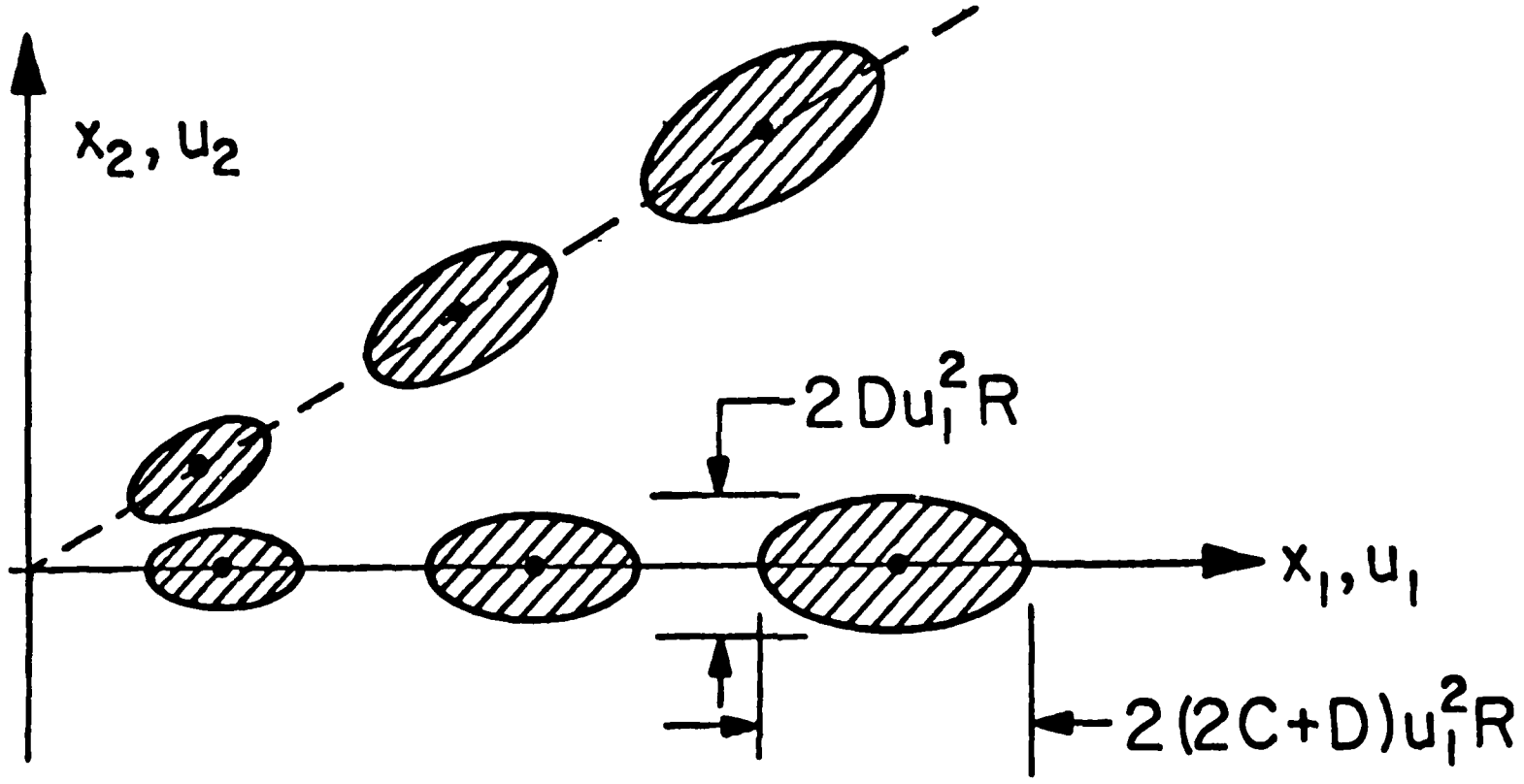


Figure 4-2. Space-variant point-spread function for astigmatism and curvature-of-field.

of the response are a function only of radial distance. Thus, the PSF $h(\cdot)$ of Eq. (4.33) written with $\eta = 0$ is just rotated about the origin to obtain the general response. Because of the inherent circular symmetry, the system complexity can be reduced by a polar coordinate transformation of the form

$$\xi = \rho_0 \cos \varphi_0 \tag{4.34a}$$

$$\eta = \rho_0 \sin \varphi_0 \tag{4.34b}$$

in both object (with subscript 0) and image (without subscript) coordinates, and rewriting Eq. (4.32) in the form

$$\rho - \rho_0 = (2C+D)\rho_0^2 r \cos \theta \tag{4.35a}$$

$$\varphi - \varphi_0 = \tan^{-1}\{D\rho_0 r \sin \theta / [1+(2C+D)\rho_0 r \cos \theta]\} = u(\rho_0) \tag{4.35b}$$

where (ρ_0, φ_0) and (ρ, φ) are the object and image polar coordinate variables. In this form, the two-dimensional space-variant radial blur becomes decoupled from the angular blur because Eq. (4.35a) does not contain φ or φ_0. The blur in the angular direction is space-invariant in φ and a slowly varying function of position ρ_0 as expressed by $u(\rho_0)$ in Eq. (4.35b).

When the degradation is purely astigmatic with no curvature-of-field, the D coefficient in Eq. (4.32) becomes zero and Eq. (4.33)

becomes singular because no blurring occurs in the angular direction. To find the SVPSF for astigmatism only, Eq. (4.33) is first collapsed to a purely radial space-variant blur by $h_\alpha(x;\xi,\eta=0)$ by evaluating

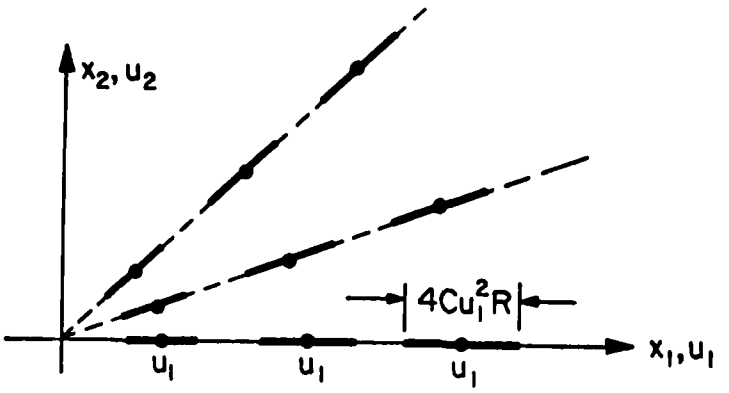
$$h_\alpha(x;\xi,\eta=0) = \int_{-\infty}^{\infty} h(x,y;\xi,\eta=0)dy \qquad (4.36)$$

and taking the limit as D approaches zero. The result is

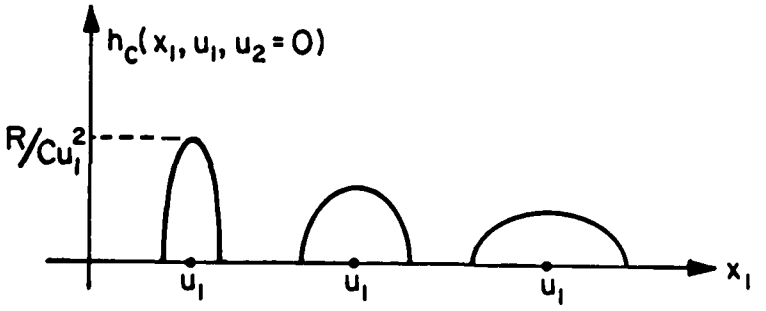
$$h_\alpha(x;\xi,\eta=0) = \frac{\left[4C^2R^2\xi^4-(x-\xi)^2\right]^{\frac{1}{2}}}{2C^2\xi^4} \quad , \quad \xi-2CR\xi^2 \le x < \xi+2CR\xi^2 \qquad (4.37)$$

and zero elsewhere. The region of support and cross section of this function are shown in Fig. 4-3. With astigmatism only, the degradation reduces to a two-dimensional space-variant line blur in a purely radial direction. Figure 4-4a is an aerial photograph displayed as 128 x 128 discrete picture elements after blurring by astigmatism with R = 1 and $C = 7.5 \times 10^{-4}$. Note that the blurring increases from zero on the optical axis (upper left corner) to nearly 50 picture elements in width as a function of increasing radius.

For the system degradations due entirely to astigmatism, the ideas of coordinate transformation restoration (CTR) [22] can be used with little modification. The basic idea is to reduce space-variant to space-invariant distortions by invertible coordinate transformation. The SVPSF of Eq. (4.37) can be modeled as a polar coordinate transformation on the object coordinates, followed by identical space-variant



(a) SVPSF for pure astigmatism with inputs at various distances u_1.



(b) Cross-section of astigmatism SVPSF

Figure 4-3. SVPSF and its cross section for pure astigmatism.

(a) Image degraded by astigmatism

(b) Polar transformation of (a)

(c) Restoration of (b)

(d) Restored image

Figure 4-4. Restoration of astigmatism.

radial blurring for each φ variable. A final inverse polar transformation on the image coordinates completes the model and produces the PSF shown in Fig. 4-3. Performing this decomposition reduces a four-dimensional space-variant restoration problem to a single two-dimensional problem.

Using the transformation, we can write the radial degradation in matrix form as

$$\underline{G}(\rho, \varphi) = \underline{H}(\rho, \rho_0)\underline{F}(\rho_0, \varphi) \tag{4.38}$$

where $\underline{G}$ and $\underline{F}$ are matrices representing image and object in polar coordinates and $\underline{H}$ is the blur matrix obtained by applying a mono-spline quadrature formulae, which was discussed in section 3.5, to the continuous space description of Eq. (4.37). This quadrature formulae provides a more accurate and smooth discrete approximation to the continuous representation of Eq. (4.37).

The space-variant restoration procedure (CVR) proceeds by inverting the two polar coordinate distortions and solving Eq. (3.38). Unfortunately, a direct inversion of $\underline{H}$ is usually not possible because point-spread function matrix $\underline{H}$ tends to be ill-conditioned leading to numerical problems. The ill-conditioning is a result of the information loss associated with the imaging process; thus H is generally singular and pseudo-inversion must be used. For inversion of Eq. (4.38), the singular value decomposition algorithm, which was discussed in the previous section, is used to obtain a unique pseudo-

inverse $\underline{H}^{+}$ which is then used in the restoration operation

$$\underline{\hat{F}}(\rho_0, \varphi) = \underline{H}^{+}(\rho_0, \rho)\underline{G}(\rho, \varphi) \tag{4.39}$$

The CTR procedure has been implemented on the image degraded only by astigmatism in Fig. 4-4a. First a polar coordinate transformation is performed to produce Fig. 4-4b in which the space-variant blur (4.38) occurs in only the radial direction. Figure 4-4c shows the restoration by SVD, in which the 7 singular values out of 128 whose magnitudes were less than 10^{-5} were not used. Figure 4-5 shows the singular values in a decreasing order. Following restoration of each line in the (ρ, φ) system, an inverse polar coordinate transformation is used to produce the final result of Fig. 4-4d.

This procedure can also be used for restoration with both astigmatism and curvature-of-field present. First, the imaging equation is expressed with a polar coordinate transformation (3.34) in the form

$$g(\rho, \varphi) = \int\int_{-\infty}^{\infty} h(\rho, \varphi; \rho_0, \varphi_0) f(\rho_0, \varphi_0) d\rho_0 d\varphi_0 \tag{4.40}$$

and then rewritten as

$$g(\rho, \varphi) = \int\int_{-\infty}^{\infty} h'(\rho, \rho_0, \varphi - \varphi_0, u(\rho_0)) f(\rho_0, \varphi_0) d\rho_0 d\varphi_0 \tag{4.41}$$

using $\varphi - \varphi_0$ and $u(\rho_0)$ of Eq. (4.35) to emphasize the functional dependence. Defining a Fourier transform of $g(\rho, \varphi)$ in the φ variable by

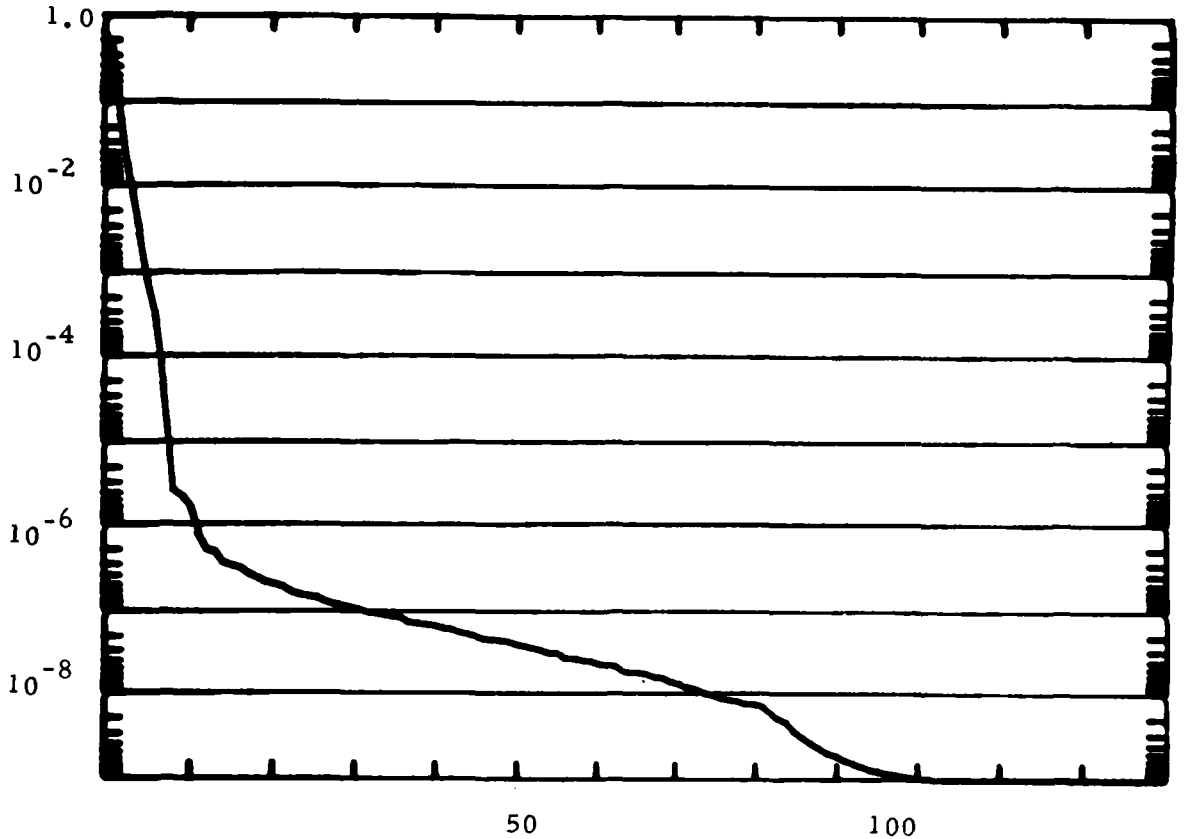


Figure 4-5. Singular values for astigmatism. $C = 7.5 \times 10^{-4}$, $R = 1$.

$$\bar{g}(\rho,\lambda) = \int_{-\infty}^{\infty} g(\rho,\varphi)\exp(-j2\pi\lambda\varphi)d\varphi \tag{4.42}$$

the transform of both sides of Eq. (4.41) is taken to obtain

$$\bar{g}(\rho,\lambda) = \iint_{-\infty}^{\infty} f(\rho_0,\varphi_0)\bar{h}(\rho,\rho_0,\lambda,u(\rho_0))\exp(-j2\pi\lambda\varphi_0)d\varphi_0 d\rho_0 \tag{4.43}$$

where $\bar{h}$ is the Fourier transform of h' in φ. Grouping terms containing φ_0 on the right side of Eq. (4.43) enables a transform in this variable to be evaluated. The resulting transformed function $\bar{f}(\rho,\lambda)$ is given by

$$\bar{f}(\rho_0,\lambda) = \int_{-\infty}^{\infty} f(\rho_0,\varphi_0)\exp(-j2\pi\lambda\varphi_0)d\varphi_0 \tag{4.44}$$

and the reduced system equation obtained from Eq. (4.41) is

$$\bar{g}(\rho,\lambda) = \int_{-\infty}^{\infty} \tilde{h}(\rho,\rho_0,\lambda)\bar{f}(\rho_0,\lambda)d\rho_0 \tag{4.45}$$

where

$$\tilde{h}(\rho,\rho_0,\lambda) = \bar{h}(\rho,\rho_0,\lambda,u(\rho_0))$$

is written as a function of three variables to show explicit dependence.

This procedure can be extended for the restoration of images degraded by simultaneous astigmatism and curvature-of-field aberrations. Following a polar coordinate transformation, a Fourier transform in φ as expressed by Eq. (4.42) is performed to partially decouple a blur as a slowly varying function of $u(\rho_0)$. The reduced

system given by the continuous space integral Eq. (4.45) has the same form as the discrete equation (4.38) for astigmatism. An estimate $\bar{f}(\rho_0, \lambda)$ is then produced by the SVD for each separate λ using similar techniques. The computational effort in this operation may be reduced by using the known variation of $\tilde{h}(\rho, \rho_0, \lambda)$ with λ. After the entire $\bar{f}(\rho_0, \lambda)$ has been obtained a series of one-dimensional inverse Fourier transforms in φ_0 is taken to find $f(\rho_0, \varphi_0)$, and an inverse polar coordinate distortion is used to get $f(\xi, \eta)$ as the final restored object. This procedure, while requiring large capabilities in computing and storage, is the only practical method for restoration of images of even moderate size. The general four-dimensional space-variant blur is effectively reduced to a set of space-variant two-dimensional problems whose point-spread functions depend in a well-known way on λ.

4.5 Overdetermined and Underdetermined Models

When the degradation matrix $\underline{H}$ of an imaging system is of full column rank, the model is called overdetermined. In practice, this usually occurs in two situations. The first is when the object has zero background and the object and image are sampled at the same rate. The second occurs when the image is sampled at a higher rate than the object.

Suppose the object function f has zero background, i.e., $f(x)$ is zero if $x < x_1$ ir $x > x_N$, and the point-spread function $h(x)$ is space-

invariant, symmetric and space-limited of width 2ℓ. Then the limits for the convolutional integral are

$$a = \max(x-\ell, x_1) \tag{4.46a}$$

$$b = \min(x+\ell, x_N) \tag{4.46b}$$

and the degraded image is given by

$$g(x) = \int_a^b h(x-\xi)f(\xi)d\xi \ . \tag{4.47}$$

Assuming uniform sampling of image, object and PSF with sampling interval $\Delta x = 1$, the continuous model can be discretized as

$$g(i) = \sum_{j=K_1}^{K_2} c_{i-j} h(i-j)f(j) \tag{4.48}$$

where c_{i-j} is the quadrature coefficient associated with h(i-j), and

$$K_1 = \max(i-L, 1) \tag{4.49a}$$

$$K_2 = \min(i+L, N) \tag{4.49b}$$

where L is the integer part of ℓ. The image sample g(i) is not zero if $-L+1 \le i \le N+L$, and thus the number of observations is

$$M = N + L - (-L+1) + 1 = N + 2L \tag{4.50}$$

Using a vector space notation, the equation (4.48) becomes

$$\underline{g} = \underline{H}\,\underline{f} \tag{4.51}$$

where H is the overdetermined M x N blur matrix defined by

$$\underline{H} = \begin{bmatrix} c_L h_L & & & 0 & & \\ \vdots & \ddots & & & \ddots & \\ c_1 h_1 & & c_L h_L & & & \\ c_0 h_0 & & \vdots & \ddots & & \\ c_1 h_1 & & c_1 h_1 & & c_L h_L & \\ \vdots & & c_0 h_0 & & \vdots & \ddots \\ c_L h_L & & c_1 h_1 & & & \\ & & \vdots & & & \\ & 0 & & & & \\ & & c_L h_L & & & \end{bmatrix} \qquad (4.51)$$

where

$$\underline{h} = (h_L, \ldots, h_1, h_0, h_1, \ldots, h_L) \qquad (4.53)$$

is the impulse response vector.

Since the vector $\underline{g}$ is in the range of $\underline{H}$, $\mathcal{R}(\underline{H})$, there exists at least one vector $\hat{\underline{f}}$ that satisfies (4.51). This estimate is unique, otherwise, if distinct vectors $\hat{\underline{f}}_1, \hat{\underline{f}}_2$ satisfy (4.51), then

$$\underline{H}\,\hat{\underline{f}}_1 - \underline{H}\,\hat{\underline{f}}_2 = \underline{H}\,\underline{g} - \underline{H}\,\underline{g} = \underline{0}$$

and

$$\underline{H}(\hat{\underline{f}}_1 - \hat{\underline{f}}_2) = \underline{0} \qquad (4.54)$$

since $\hat{\underline{f}}_1 - \hat{\underline{f}}_2 \neq 0$, Eq. (4.54) indicates that a linear combination of columns of $\underline{H}$ is zero which contradicts the assumption of an over-determined model. This unique solution can be obtained by the pseudoinverse of a full rank matrix of by SVD algorithm discussed in section 4.3. In noisy images, $\underline{g}$ is not necessarily in $\mathcal{R}(\underline{H})$, and a least square criteria based on (4.11) leads to a unique solution [16].

A more realistic model for an imaging system can be obtained if no restrictions are imposed on the background of the object. In practice, few objects are recorded with a background of zero or known intensity. Moreover, because of computational problems, an image is often partitioned into sections before being processed, and thus the assumption of zero background cannot be valid. A model must be used that relates a portion of the object to the corresponding segment of the image without any restrictions on the background. In such a model, because of blurring, a portion of the image is affected by a larger segment of the object. Therefore, if the object and image are sampled with the same rate, the matrix $\underline{H}$ has more columns than the rows, and the system is underdetermined. Following the same procedure as with overdetermined models, the system is represented by

$$\underline{g} = \underline{H}\,\underline{f} \tag{4.55}$$

where $\underline{H}$ is M x N blur matrix, and

$$M = N - 2L$$

The matrix $\underline{H}$ is given by

$$\underline{H} = \begin{bmatrix} c_L h_L & \cdots & c_1 h_1 & c_0 h_0 & c_1 h_1 & \cdots & c_L h_L & & \\ & \ddots & & \ddots & & \ddots & & \ddots & \\ 0 & & \ddots & & \ddots & & \ddots & & 0 \\ & & & c_L h_L & \cdots & c_1 h_1 & c_0 h_0 & c_1 h_1 & \cdots & c_L h_L \end{bmatrix} \qquad (4.56)$$

where $\underline{h}$ and $\underline{c}$ are the same as overdetermined model. As mentioned in section 4.3, the system (4.55) does not have a unique solution. The minimum norm solution based on (4.10) and (4.11) is unique and can be obtained by (4.26) if $\underline{H}$ is of full row rank, or by the SVD algorithm.

4.6 Experimental Results

To illustrate image restoration by pseudo-inversion, Fig. 4-6a is selected as a test scene which is originally of size 128 x 128 picture elements (pixels) with 110 x 110 nonzero elements. For display, this zero background picture has been enlarged by cubic spline interpolation to an image of size 256 x 256 pixels. Figure 4-6b represents the image after undergoing motion degradation with a blurring point spread function of length 11 pixels. This is a severe blur because it is 1/10 of the original picture size. The observed image is of

(a) Original

(b) Blurred

(c) Restored

Figure 4-6. Restoration of motion degradation, overdetermined model.

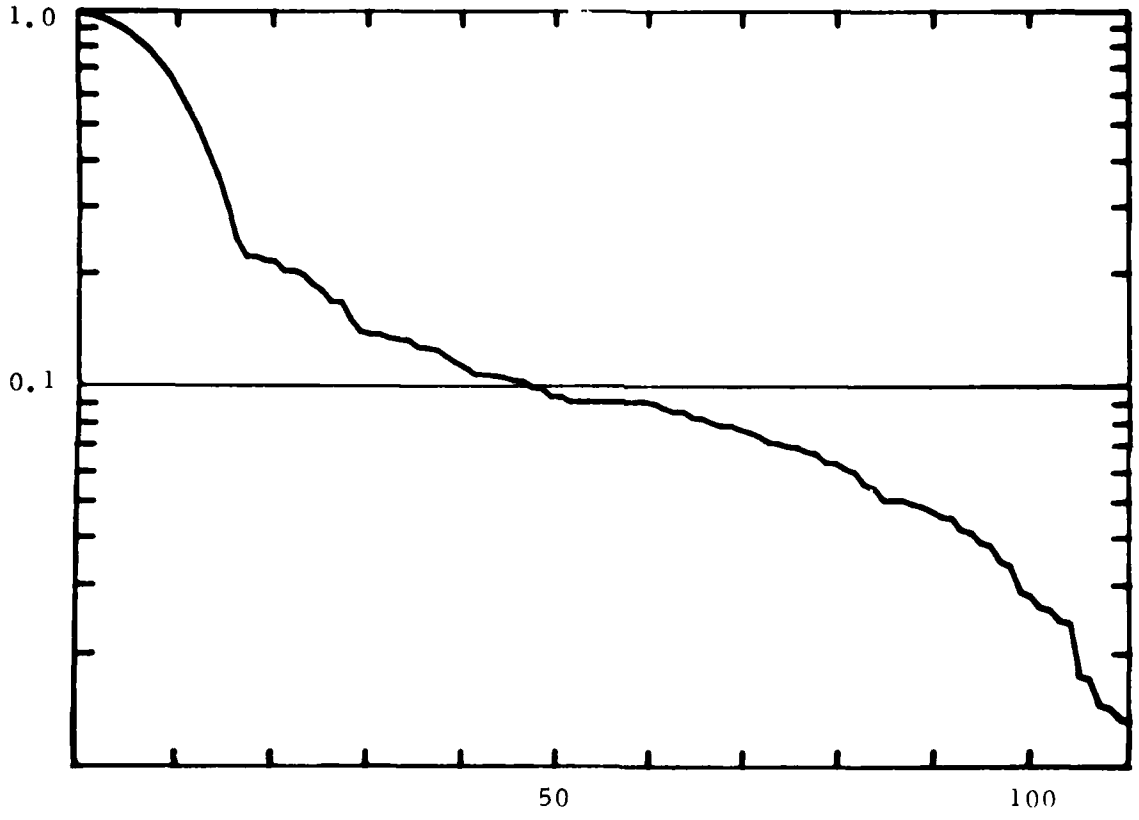


Figure 4-7. Singular values for motion blur.

size 110 x 120 and thus the system is overdetermined as discussed in section 4.5. The blur matrix $\underline{H}$ is a 120 x 110 banded matrix given by Eq. (4.52) with 11 nonzero elements on each column. Although this matrix is of full column rank, because of high dimensionality and ill-conditioning, using Eq. (4.27) and inverting $\underline{H}^t\underline{H}$ produces computational difficulties. Thus, the SVD algorithm has been employed to compute the pseudo-inverse of matrix $\underline{H}$. Figure 4-7 shows the singular values of $\underline{H}$ ordered in decreasing value. Since the singular values are much greater than the machine precision and there is not a sudden decrease in value, all of them have been used used for computation of $\underline{H}^+$. Figure 4-6c shows the restored image. As far as visual perception is concerned, this restored image is identical to the original scene.

Chapter 5

RESTORATION OF NOISY IMAGES

In the previous chapter, the problem of image degradation and restoration was modeled for a noiseless system. In reality, an image is often affected by a variety of noisy sources. The scanner measurement error is a source of noise which adds some element of uncertainty to the measured signal. The quantizer, that maps the signal amplitude to a finite number of digital levels for computational and coding purposes, is another source of noise. Coding and channel errors occur when the image is transmitted through a noisy channel. Computers produce truncation and roundoff errors due to the limits of machine precision. Photographic film is the most common recording system in image processing and is another source of noise which will be discussed in more detail in section 5.3.

Although, these examples are not an exhaustive list of noise sources, noise of various types is a common problem in every type of imaging system. In this chapter the effects of noise in image restoration is considered from several viewpoints.

5.1 Discrete Wiener Filter

The basic idea of the Wiener filter and its derivation in a continuous model was discussed in section 2.3. In this section, the same fidelity criteria is applied to a discrete model and the

filter is obtained.

In the presence of noise, the discrete imaging model is given by [49]

$$\underline{g} = \underline{H}\,\underline{f} + \underline{n} \tag{5.1}$$

where $\underline{H}$ is the blur matrix and $\underline{n}$ is the noise vector. Assuming the vectors $\underline{f}$ and $\underline{n}$ are members of two random processes, the minimum mean square estimate of $\underline{f}$ is $\hat{\underline{f}}$ such that it minimizes the objective function

$$e = E[(\underline{f} - \hat{\underline{f}})^2] \tag{5.2}$$

where E denotes the expectation over processes $\underline{f}$ and $\underline{n}$. Let the estimate $\hat{\underline{f}}$ be represented by

$$\hat{\underline{f}} = \underline{R}\,\underline{g} \tag{5.3}$$

where $\underline{R}$ is a linear filter. Using the orthogonality principle [50], the observation $\underline{g}$ and the error $\underline{f} - \hat{\underline{f}}$ must satisfy the following relation

$$E[(\underline{f} - \hat{\underline{f}})\underline{g}^t] = \underline{0} \tag{5.4}$$

Substituting (5.1) and (5.3) in Eq. (5.4), and assuming $\underline{f}$ and $\underline{n}$ to be zero-mean uncorrelated random vectors, the Wiener filter is derived as

$$\underline{R} = \underline{C_f}\underline{H}^t(\underline{H}\,\underline{C_f}\underline{H}^t + \underline{C_n})^{-1} \tag{5.5}$$

where $\underline{C_f}$ and $\underline{C_n}$ are the correlation matrices of the signal and noise, respectively. Clearly, it is assumed that $\underline{H}\,\underline{C_f}\underline{H}^t + \underline{C_n}$ is non-

singular. If the noise is white, this condition is satisfied, because $\underline{H}\,\underline{C}_f\underline{H}^t$ is non-negative-definite and $\underline{C}_n$ is positive-definite, therefore $\underline{H}\,\underline{C}_f\underline{H}^t + \underline{C}_n$ is positive-definite and consequently is nonsingular. Another version of (5.5) can be obtained by using the matrix identity [51]

$$\underline{A}\,\underline{B}^t(\underline{C} + \underline{B}\,\underline{A}\,\underline{B}^t)^{-1} = (\underline{A}^{-1} + \underline{B}^t\underline{C}^{-1}\underline{B})^{-1}\underline{B}^t\underline{C}^{-1} \tag{5.6}$$

in Eq. (5.5)

$$\underline{R} = (\underline{H}^t\underline{C}_n^{-1}\underline{H} + \underline{C}_f^{-1})\underline{H}^t\underline{C}_n^{-1} \tag{5.7}$$

where $\underline{C}_n$ and $\underline{C}_f$ are assumed to be nonsingular. Although these two versions are equivalent, their computation in practice depends on the structure of $\underline{H}$, $\underline{C}_n$ and $\underline{C}_f$. If $\underline{H}$ has fewer columns than rows, $\underline{H}\,\underline{C}_f\underline{H}^t + \underline{C}_n$ has a higher dimension than $\underline{H}^t\underline{C}_n^{-1}\underline{H} + \underline{C}_f^{-1}$, and therefore needs more computations for inversion. The order is reversed when the number of columns is more than the number of rows. Moreover, Eq. (5.7) requires the inversion of two more matrices; $\underline{C}_n$ and $\underline{C}_f$.

The Wiener filter can also be obtained by minimization of the objective function

$$W(\underline{f}) = \|\underline{C}_f^{-\frac{1}{2}}\underline{f}\|^2 + \|\underline{C}_n^{-\frac{1}{2}}(\underline{g}-\underline{H}\,\underline{f})\|^2 \tag{5.8}$$

where $\underline{C}_f^{\frac{1}{2}}$ and $\underline{C}_n^{-\frac{1}{2}}$ are the whitening filters for the signal and noise, respectively. Taking derivatives with respect to $\underline{f}$

$$\frac{dW}{d\underline{f}} = \underline{C}_f^{-1}\underline{f} + \underline{H}^t\underline{C}_n^{-1}\underline{H}\underline{f} - \underline{H}^t\underline{C}_n^{-1}\underline{g}$$

and setting the result equal to zero, the estimate is

$$\hat{\underline{f}} = (\underline{H}^t\underline{C}_n^{-1}\underline{H} + \underline{C}_f^{-1})^{-1}\underline{H}^t\underline{C}_n^{-1}\underline{g} \qquad (5.9)$$

which is the same as (5.7). The inversion of large size matrices in Eqs. (5.5) and (5.7) is not an easy task, particularly because of ill-conditioning. Using the Fourier transform for the inversion of circulant matrices leads to a fast Wiener filter with much less computation and higher efficiency [13].

5.2 Filtering of Unblurred Noisy Images by Smoothing Spline Functions

As discussed in Chapter 1, the Wiener filter has many limitations and shortcomings. Not only does it require the most a priori information, it often produces an estimate of the restored image with poor visual quality in comparison to other filters [52]. The constrained least squares estimate which has been introduced as an alternative to the Wiener filter is capable of producing images with much higher visual quality. Moreover, this estimate requires less a priori information than the Wiener filter [20].

In this section, a constrained least squares estimate based on the minimization of the second derivative and local statistics of the noise is defined and the corresponding filter, which produces a cubic spline function as the estimate, is obtained. The parameters of this filter

determine the local smoothing window and overall extent of smoothing, and thus one can control the tradeoff between resolution and smoothing in a spatially nonstationary manner. Since the derivation of the filter for unblurred noisy images is simpler and the estimate can be computed by highly efficient algorithms, the derivation of the filter for unblurred noisy images is covered here. The corresponding filter for noisy blurred images will be covered in section 5.5.

For an unblurred noisy image, the image $g(x)$ is given by

$$g(x) = f(x) + n(x) \tag{5.10}$$

in a continuous model. The additive noise $n(x)$ is assumed to be an uncorrelated zero mean random process, i.e.,

$$E[n(x)] = 0, \quad E[n(x_1)n(x_2)] = 0 \quad \text{for } x_1 \neq x_2. \tag{5.11}$$

In order to filter the image, the following fidelity criteria is defined:

$$\text{minimize} \int (f''(x))^2 dx \tag{5.12}$$

among all twice differentiable functions $f \in C^2$ such that

$$\sum_i \left(\frac{g(x_i)-f(x_i)}{\delta_i} \right)^2 \leq S \tag{5.13}$$

where the positive number δ_i locally controls the smoothing window at point x_i and S controls the overall extent of smoothing. If σ_i, the standard deviation of noise or its estimate at point x_i is available, then it can be used for δ_i. In this case, natural values of S lie within the

confidence interval of the left side of (5.13) that is

$$N - (2N)^{\frac{1}{2}} \leq S \leq N + (2N)^{\frac{1}{2}} \qquad (5.14)$$

where N is the number of data points. Reinsch [53] has shown that the solution to (5.12) and (5.13) is a cubic spline and more generally is a spline function of degree 2K-1 for least squares minimization of the K^{th} derivative instead of the second derivative [54]. The case K = 2 leads to very simple algorithms for the construction of f(x). Moreover, cubic splines give satisfactory results and are easy to evaluate. Choosing S equal to zero implies

$$f(x_i) = g(x_i) \qquad i = 1, \ldots, N$$

which leads to the problem of interpolation by cubic spline functions.

Applying the well-known Lagrange multiplier method, along with the auxiliary variable z to change the inequality constraint (5.13) to an equality constraint, the objective function

$$\int_{x_1}^{x_N} (f''(x))^2 dx + p\left[\sum_{i=1}^{N} \left(\frac{g(x_i)-f(x_i)}{\delta_i}\right)^2 + z^2 - S\right] \qquad (5.15)$$

must be minimized. The optimal function f(x) is determined as the following

$$f(x_i)_- - f(x_i)_+ = 0 \qquad i = 1, \ldots, N \qquad (5.16a)$$

$$f'(x_i)_- - f'(x_i)_+ = 0 \qquad i = 1, \ldots, N \qquad (5.16b)$$

$$f''(x_i)_- - f''(x_i)_+ = 0 \qquad i = 1, \ldots, N \qquad (5.16c)$$

$$f'''(x_i)_- - f'''(x_i)_+ = 2p\frac{f(x_i)-g(x_i)}{\delta_i^2} \qquad i = 1,\ldots,N \tag{5.16d}$$

where $f^{(k)}(x)_\pm = \lim_{h\to 0} f^{(k)}(x\pm h)$ for K = 0,...,3. Moreover

$$f''''(x) = 0, \quad x_i < x < x_{i+1}, \quad i = 1,\ldots,N-1. \tag{5.16e}$$

Equations (5.16) indicate that the function f(x) is composed of piecewise polynomials of degree 3 or less in each interval $[x_i, x_{i+1}]$ of the form

$$f(x) = a_i + b_i(x-x_i) + c_i(x-x_i)^2 + d_i(x-x_i)^3 \tag{5.17}$$

such that they are continuous up to the second derivative at their joining points, and thus the solution is a cubic spline. Assuming f(x) to be a natural spline [33] of degree 3, the following extra conditions

$$f''(x_1)_- = f'''(x_1)_- = f''(x_N)_+ = f'''(x_N)_+ = 0 \tag{5.18}$$

must be satisfied. Substituting (5.17) in (5.16) and (5.18), the optimal filter with respect to conditions (5.12) and (5.13) is obtained as [Appendix A]

$$\hat{\underline{f}} = \underline{g} - \underline{D}^2\underline{Q}(\underline{Q}^t\underline{D}^2\underline{Q} + p\underline{\Gamma})^{-1}\underline{Q}^t\underline{g} \tag{5.19}$$

where

$$\underline{f} = [f(x_1), f(x_2), \ldots, f(x_N)]^t \tag{5.20a}$$

$$\underline{g} = [g(x_1), g(x_2), \ldots, g(X_N)]^t \tag{5.20b}$$

$$\underline{D} = \text{diag}[\delta_1, \delta_2, \ldots, \delta_N]. \tag{5.20c}$$

Matrix $\underline{\Gamma}$ is a positive definite tridiagonal matrix of order N-2, and $\underline{Q}$ is a tridiagonal matrix with N rows and N-2 columns. For a uniformly sampled image with unit sampling interval, $\underline{Q}$ and $\underline{\Gamma}$ have the form

$$\underline{Q} = \begin{bmatrix} 1 & & & 0 \\ -2 & 1 & & \\ 1 & -2 & \ddots & \\ & 1 & & 1 \\ 0 & & & -2 \\ & & & 1 \end{bmatrix} \tag{5.21a}$$

$$\underline{\Gamma} \begin{bmatrix} 4/3 & 1/3 & & & 0 \\ 1/3 & 4/3 & 1/3 & & \\ & \ddots & \ddots & \ddots & \\ 0 & & 1/3 & 4/3 & 1/3 \\ & & & 1/3 & 4/3 \end{bmatrix} \tag{5.21b}$$

Here matrix $\underline{Q}$ represents a second order differentiation matrix. Since Q is a matrix of full column rank, and $\underline{D}$ is nonsingular, the matrix $\underline{Q}^t\underline{D}^2\underline{Q} + p\underline{\Gamma}$ has an inverse for all values of $p \geq 0$. Thus, if p is known, the estimate $\hat{\underline{f}}$ can be obtained by (5.19). The Lagrangian parameter p, like S, controls the overall extent of smoothing and

may be determined in terms of S.

The objective function (5.15) has to be minimized also with respect to z and p, leading to the conditions

$$pz = 0 \tag{5.22}$$

$$\sum_{i=1}^{N} \left(\frac{f(x_i)-g(x_i)}{\delta_i} \right)^2 = S - z^2 . \tag{5.23}$$

Substituting (5.19) in (5.23) gives

$$F(p) = \| \underline{D}\,\underline{Q}(\underline{Q}^t\underline{D}^2\underline{Q} + p\underline{\Gamma})^{-1}\underline{Q}^t\underline{g}\| = (S-z^2)^{\frac{1}{2}} \tag{5.24}$$

Condition (5.22) implies either p = 0 or z = 0. The first case is only possible if $F(0) \leq S$, and thus the straight line fitted to data points by least square principles satisfies condition (5.13) and the cubic spline reduces to this straight line. If $F(0) > S$, then $p \neq 0$ and z = 0, the inequality constraint (5.13) changes to an equality constraint and Eq. (5.24) can be written as

$$F(p) = S^{\frac{1}{2}} \tag{5.25}$$

Reinsch [40] has shown that there exists a unique solution for p satisfying (5.25) and (5.12). This positive unique solution can be determined by using Newton's method.

Since the parameters p and S are interrelated by (5.25) and both have the same effect on filtering, by selecting p instead of S, one can skip the iterative Newton's method to compute p and reduce

the numerical operations considerably.

The matrix $\underline{\Gamma}$ is positive definite and invertible, by using matrix identity [Appendix B], Eq. (5.19) can thus be written in the more concise form:

$$\hat{\underline{f}} = (I + p^{-1}\underline{D}^2\underline{Q}\,\underline{\Gamma}^{-1}\underline{Q}^t)^{-1}\underline{g} \qquad (5.26)$$

Although (5.26) appears simpler, it needs two matrix inversions in comparison to Eq. (5.19) which requires the inverse of a banded matrix. The Cholesky decomposition [55] $\underline{R}^T\underline{R}$ of a positive-definite band matrix $\underline{Q}^t\underline{D}^2\underline{Q} + p\underline{\Gamma}$, where $\underline{R}$ is a lower diagonal (triangular) matrix, provides an efficient computational algorithm for performing the filtering of Eq. (5.19).

5.3 Application of Smoothing Spline Filter in Signal-Dependent Noisy Images

An interesting property of the fidelity criterion (5.13) is that the smoothing window can be locally controlled by determination of δ_i. If the noise variance is higher in some regions, δ_i can be set larger at that region. This property enables the spline filter capable to restore images even with the difficult problems of signal-dependent or multiplicative noise. A common source of this type of noise is photographic film, and since it is widely used as a recording system in image processing, the restoration of images degraded by film-grain noise is very desirable.

Image formation on a photographic film is a complex optical and

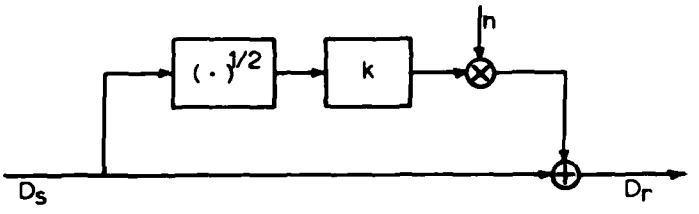
chemical process. A very accurate model of this process, even if possible, is too complicated to be used, and a very simplified model can hardly represent the actual physical process. A reasonable and fairly accurate model for film grain noise is given by [56], [58]

$$D_r = D_s + kD_s^{\frac{1}{2}}n \qquad (5.27)$$

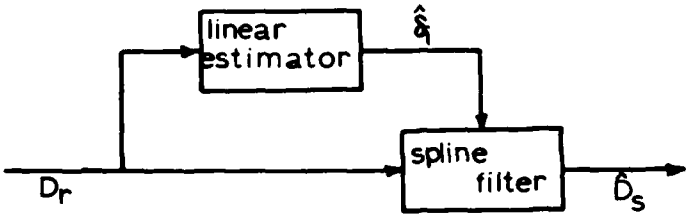
where D_s is the signal density, D_r is recorded density on film and n is zero mean Gaussian with unit variance that is statistically independent of the signal. A block diagram of this model is shown in Fig. 5-1a.

Equation (5.27) shows that when the signal has a higher amplitude, noise has a greater variance. To apply the above mentioned filter, the value of δ_i and an estimate of the signal are needed. Hunt and Cannon [59] have shown that images are well described statistically as a stationary variance about a spatially non-stationary local mean. Assuming ergodicity of similar classes of images, the spatial average may be used as an estimate of the ensemble average. A local spatial average is thus used as a non-stationary estimate of the signal mean. Since a linear estimate of δ_i in terms of the observation is used in a nonlinear manner to filter the image, the overall filtering is nonlinear. The block diagram of this filter is shown in Fig. 5-1b.

The filter of Eq. (5.19) has been applied to an image corrupted

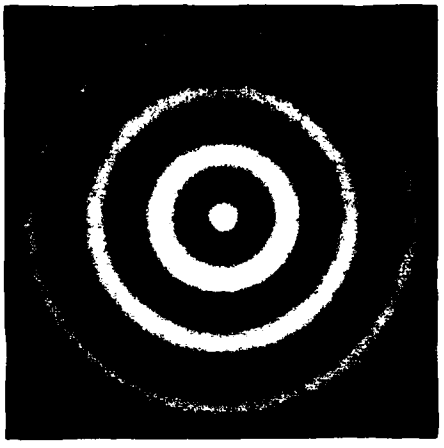


(a) Block diagram of film-grain noise



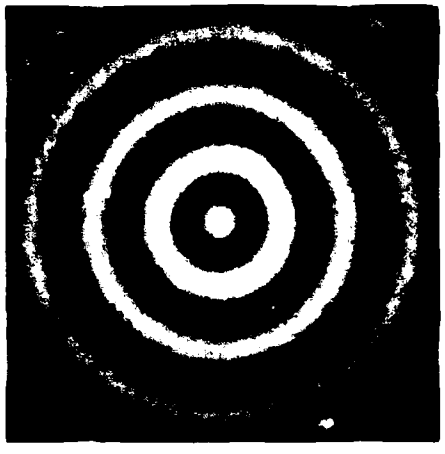
(b) Filtering of film-grain noise

Figure 5-1. Models for film-grain noise and filtering.



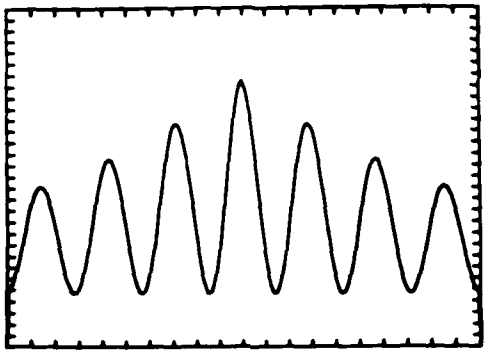
(a) Original

(b) Noisy, signal-dependent

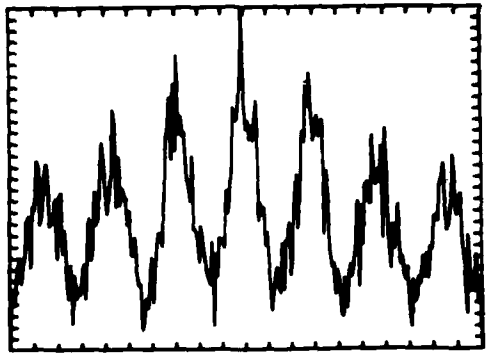


(c) Filtered

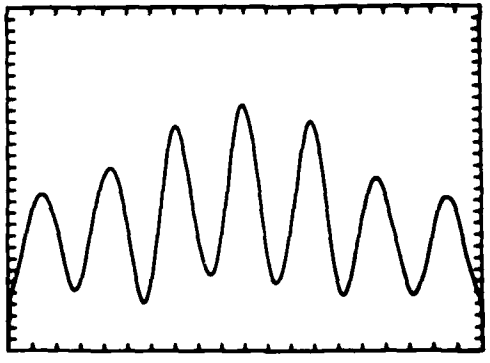
Figure 5 2. Filtering of signal dependent noisy images.



(a) Original



(b) Noisy, film-grain noise



(c) Filtered

Figure 5-3. Filtering of image lines degraded by film-grain noise.

by signal-dependent noise modeled by Eq. (5.27) when k = 2. Figure 5-2a is the original image, Fig. 5-2b is the noisy image and Fig. 5-2c is the filtered result. Figure 5-3a is a plot of the brightness cross-section of Fig. 5-2a, and the subsequent plots are brightness cross-sections of noisy and filtered images. The filter has reduced the mean square error by a factor of ten.

5.4 Restoration of Noisy Blurred Images

A noisy blurred image can be expressed in the form

$$\underline{g} = \underline{H}\,\underline{f} + \underline{n} \tag{5.28}$$

for a discrete model, where $\underline{f}$ and $\underline{g}$ are samples of the object and image functions given by (5.20a) and (5.20b), respectively. The following fidelity criteria for image restoration may then be formulated as minimization of

$$\int_{x_1}^{x_N} (f''(x))^2 dx \tag{5.29}$$

among all twice differentiable function $f \in C^2$ such that

$$\| \underline{D}^{-1}(\underline{H}\,\underline{f} - \underline{g}) \|^2 \leq S \tag{5.30}$$

where $\underline{D}$ is defined by Eq. (5.20c). This minimization criteria is similar to what was defined for unblurred noisy images with a slight modification. If $\underline{H} = I$, the condition (5.30) reduces to (5.13). Using the same procedure as before, the objective function

$$\int_{x_1}^{x_N} (f''(x))^2 dx + p[\|\underline{D}^{-1}(\underline{H}\,\underline{f}-\underline{g})\|^2 + z^2 - S] \qquad (5.31)$$

must be minimized with respect to $\underline{f}$. Assuming the function $\underline{f}$ to be composed of piecewise polynomials given by (5.17), the equations

$$\underline{\Gamma c} = \underline{Q}^t \underline{a} \qquad (5.32)$$

$$\underline{Q}\,\underline{c} = p\underline{H}^t\underline{D}^{-2}(\underline{g}-\underline{H}\,\underline{a}) \qquad (5.33)$$

are obtained, where

$$\underline{c} = [c_2, c_3, \ldots, c_{N-1}]^t \qquad (5.34)$$

$$\underline{a} = [a_1, a_2, \ldots, a_N]^t \qquad (5.35)$$

and $\underline{Q}$ and $\underline{\Gamma}$ are previously defined. In order to satisfy condition (5.18), c_1 and c_n are assumed to be zero. Since $\underline{\Gamma}$ is a positive definite matrix

$$\underline{c} = \underline{\Gamma}^{-1}\underline{Q}^t\underline{a} \qquad (5.36)$$

and substituting (5.36) in (5.33) yields

$$\underline{Q}\,\underline{\Gamma}^{-1}\underline{Q}^t\underline{a} = p\underline{H}^t\underline{D}^{-2}(\underline{g}-\underline{H}\,\underline{a})\,. \qquad (5.37)$$

From (5.17), it is clear that $\underline{a} = \hat{\underline{f}}$, therefore

$$\hat{\underline{f}} = (\underline{H}^t\underline{D}^{-2}\underline{H} + \lambda\underline{Q}\,\underline{\Gamma}^{-1}\underline{Q}^t)^{-1}\underline{H}^t\underline{D}^{-2}\underline{y} \qquad (5.38)$$

where $\lambda = p^{-1}$. Setting $\underline{H} = I$ leads to the estimate of Eq. (5.26). Comparison of this restoration filter to the Wiener filter given by (5.7) shows the similarity of these two filters with $\underline{C}_f^{-1}$ replaced by

$\lambda \underline{Q}\ \underline{T}^{-1}\underline{Q}^{t}$. In order to obtain the filter numerically, it is necessary to invert the matrix $\underline{T}$ as well as $\underline{H}^{t}\underline{D}^{-2}\underline{H} + \lambda \underline{Q}\underline{T}^{-1}\underline{Q}^{t}$. Matrix $\underline{T}$ is a positive-definite tridiagonal matrix and efficient techniques exist for the computation of its inverse [57]. Since $\underline{H}$ is a large matrix for ordinary size images, and also H is not of full column rank for non-zero background pictures, the matrix $\underline{H}^{t}\underline{D}^{-2}\underline{H} + \lambda \underline{R}\ \underline{\Gamma}^{-1}\underline{Q}^{t}$ may be ill-conditioned or singular and thus pseudo-inversion is used instead of exact inversion. In section 5.6 the experimental results of this filter are presented.

5.5 The Effect of Fidelity Criteria on High Frequency Suppression

In Chapter 4, describing the restoration of noiseless images, a fidelity criteria based on the minimization of the L_2-norm of the function was defined which led to the filter given by (4.19) and pseudoinversion. In this chapter, the fidelity criterion is based on the minimization of L_2-norm of the second derivative of the function, leading to spline filtering and restoration. In this section, the effects of different criteria on the frequency content of the restored image are discussed.

The objective function for both cases has the one common term

$$\lambda \int (f^{(k)}(x))^2 dx \qquad (5.39)$$

where λ is a weighting factor. If k = 0, the derived filter is given by (4.19) or by pseudo-inversion, and k = 2 leads to spline restoration.

Minimization of the objective function implies the reduction of term (5.39) in addition to the other terms. Taking the Fourier transform of this term and applying Parseval's theorem, it is equivalent to

$$\lambda \int [(2\pi u)^k F(u)]^2 du \qquad (5.40)$$

where u is the spatial frequency and F(u) is the Fourier transform of f(x). As (5.40) shows, the amplitude F(u) at spatial frequency u is weighted by a factor $(2\pi u)^k$ and then squared and integrated. Therefore, as k increases, the weighting factor of high frequency amplitude increases too. Thus, reducing (5.40) with larger k implies the suppression of the higher frequencies more than the lower frequencies. When k = 0, the amplitudes of all the frequencies are equally weighted.

In most image restoration problems with white noise, the original object is a relatively low frequency signal, and noise has a flat power spectral density. When the noise power is zero or very low, there is no need to suppress the high frequency content of the image. Moreover, preserving the higher frequencies preserves the fine details of the image, and thus, k = 0 is a good choice for noiseless images. Conversely, for noisy images, where most of the high frequency content of the image belongs to the noise, k = 2 is a good choice. Generally, as k increases, the filtered signal is more smooth and correlated with less higher frequency content.

5.6 Experimental Results

In this section the experimental results of restoration of noisy blurred images by spline functions are presented. Figure 5-4a illustrates a scene of original size 128 x 128 pixels enlarged to 256 x 256 by cubic spline interpolation for display purposes. Figure 5-4b represents the image following motion degradation with a point-spread function of length 7 pixels and the addition of zero mean Gaussian noise with standard deviation 1. The observed image has a nonzero background, and thus the system is underdetermined. For restoration, the smoothing parameter λ of the filter (5.38) is set to .01 and .001, and the SVD algorithm is employed to compute the pseudo-inverse of the matrix $\underline{H}^t\underline{D}^{-2}\underline{H} + \lambda\underline{Q}\,\underline{\Gamma}^{-1}\underline{Q}^t$.

Figure 5-5 shows the behavior of singular values for different values of λ. When λ is small, it only affects the smaller singular values (right side portion of the curves) which correspond to high frequency eigenvectors, while for large λ, the smoothing term $\underline{Q}\,\underline{\Gamma}^{-1}\underline{Q}^t$ completely dominates the deblurring term $\underline{H}^t\underline{D}^{-2}\underline{H}$ and its singular values. The effect of λ extends from the left side portion of the curves to the right portion as λ increases. For the computation of the pseudoinverse a judicious threshold ϵ must be selected for the cutoff of singular values. Figures 5-4c and 5-4d show the restored image for $\lambda = .01$, $\epsilon = .005$, and for $\lambda = .001$, $\epsilon = .001$, respectively. As λ and ϵ decrease, the restored image becomes sharper with

(a) Original

(b) Noisy blurred, L=7, σ=1

(c) Restored. $\lambda = .01$, $\varepsilon = .005$

(d) Restored. $\lambda = .001$, $\varepsilon = .001$

Figure 5-4. Restoration of noisy blurred images by spline filter.

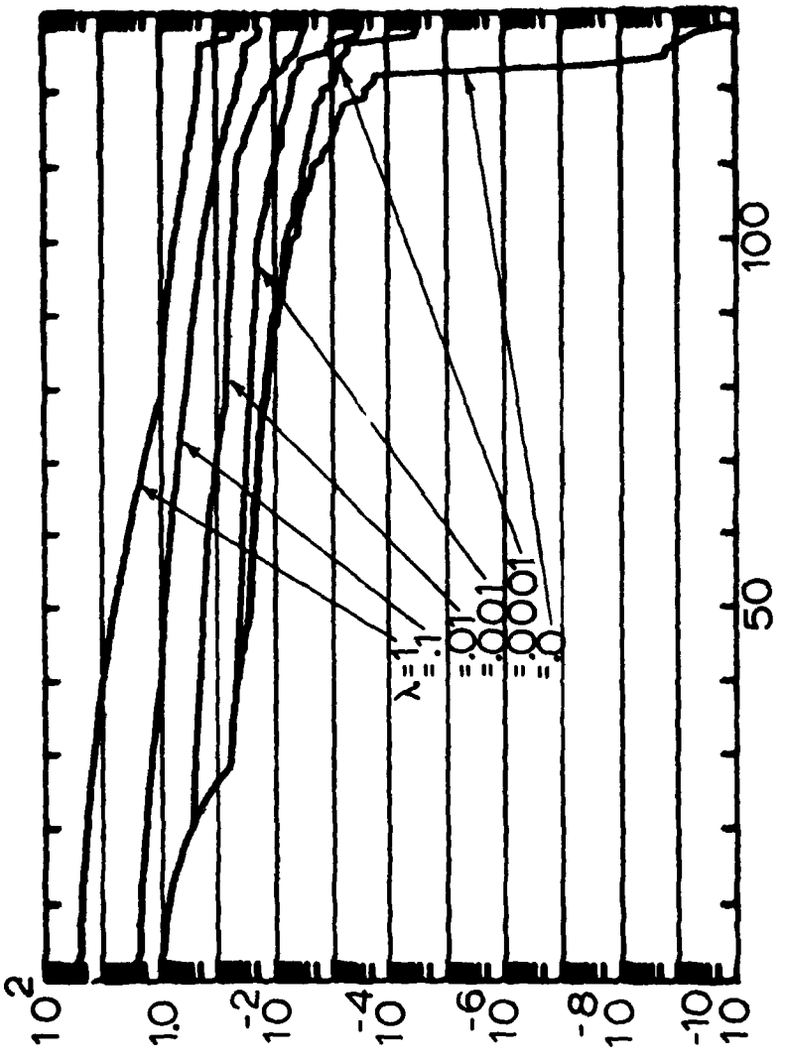


Figure 5-5. Singular values of spline filter.

more details, but at the same time the amplitude of unwanted high frequency components increases. For large λ and ϵ, the restored image is smoother and more correlated. Thus, by choosing these two parameters as well as the local smoothing window δ_i, one can control the tradeoff between smoothing and resolution locally and globally. It is obvious that the quality of the restored image is highly dependent on the proper selection of these parameters.

Chapter 6

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

This dissertation has presented a theoretical and experimental analysis of image restoration in the spline domain. Representing object, image and point-spread functions by splines has led to a more accurate and realistic model. The interesting properties of spline functions, particularly B-splines, have been used in image modeling and restoration.

The linear integral equation that describes the image formation has been discretized for processing by a digital computer. Various methods such as pulse approximation, Newton-Cotes and monospline quadrature formulae are discussed and compared with each other. The first two methods are special and extreme cases of the third one. It is shown that the monospline quadrature formulae of degree 3 produces less error overall than the other two methods, although the best choice of degree for the monospline is dependent on the variation of the integrand and sampling interval. For the class of rapidly varying functions, a smaller m is advised, but for the class of slowly varying functions, large values of m give better results. These results are true for both undersampled and oversampled functions.

B-splines have been used for interpolation and approximation of object and image functions. B-splines have some advantages over the sine and cosine basis functions of the Fourier domain. First, they are strictly positive and thus they are a better representation of image intensity. Second, the shape of these functions is fixed except for a shift of location. This property results in circulant matrices which are easy to handle numerically. Third, their local basis properties results in banded matrices. Fourth, their convolutional property represents the convolution integral of space-invariant degradations. It has been shown that using B-splines and exploiting the convolutional property for restoration of an analytically blurred image results in a more accurate reconstruction of the original object. Since the blurred image is generally smoother than the object, a higher degree B-spline represents it better than the one representing the object. The degree of B-spline must be selected according to the frequency content (or variation) of the approximated function.

The minimum norm criteria leading to the pseudo-inversion has been used for the restoration of space-variant degradations, over-determined models and underdetermined models. A numerical technique called singular-value-decomposition is used for computation of the pseudo-inverse. Due to the singularity of the system, ill-conditioning or roundoff error, a judicious threshold for the nonzero singular values must be made. A sudden decrease in their

values and the machine precision are the determining factors. The strikingly good results for the restoration of astigmatism which is strongly space-variant show the validity of the objective criteria and the capability of the numerical technique.

A constrained least squares criteria based on the minimization of the second derivative and local statistics of the noise has been defined and the optimal filter, which produces a cubic spline function as the estimate, has been derived. This filter is applicable to both space-variant and space-invariant degradations. The parameters of the filter determine the local smoothing window and overall extent of smoothing and thus the tradeoff between resolution and smoothing is controllable both locally and globally. These properties have made the spline filter capable of restoring images with signal-dependent or multiplicative noise. This filter has been successfully applied for filtering images degraded by film-grain noise which is modeled as signal-dependent or multiplicative noise. Since a local spatial average has been used as a nonstationary estimate of the signal mean and local smoothing factor, the overall filtering is nonlinear. The filter has been also applied to images degraded by motion blur and noise. The results show that as the smoothing parameter decreases, the restored image becomes sharper with more details, but at the same time the amplitude of unwanted high frequency components increases. For large smoothing parameters, the restored images

become more correlated. Finally, the effect of fidelity criteria on high frequency components of the restored image has been analyzed. If the noise power is zero or almost zero, the fidelity criteria based on the minimization of the function is advised, but for noisy images, the fidelity criteria with respect to the minimization of the second derivative of the function gives better results.

The research pursued in this dissertation may be extended in several directions. A more detailed study of monospline quadrature formulae, particularly free nodes, would be of considerable interest. The number of nodes and their locations have substantial effect on quadrature error and accuracy of the discrete model. This study can be extended with respect to the behavior and local properties of the integrand. In approximating the image function by splines, variable sampling may be studied for optimal knot placement and error reduction. A discussion of the problem as well as some algorithms may be found in references [60], [61].

A more detailed experimental study of the spline filter is particularly useful in determining the parameters of the filter according to the local properties of the image and noise statistics. The filter can be applied to non-uniform sampling with minor modification of the filter matrices [Appendix A]. Thus the sampling intervals may be determined with respect to local behavior of the observed image and then the modified filter can be used.

Recursive computational techniques are of special interest in real-time processing of the images. A version of the filter may be developed for restoring images when the stream of data comes from a scanner or transmitter. Another possibility would be an exploration of the iterative method for solution of the filter. Such a method has been developed for computation of the pseudoinverse [62], [63]. Since the matrices of the spline filter are banded circulant or Toeplitz, very efficient numerical techniques, particularly for space-invariant degradations, may be used.

As mentioned earlier, spline functions have many desirable approximating and interpolating characteristics that can be used in various areas of digital image processing. So far, some attention has been given to potential applications of spline functions in image restoration and enlargement [45], [64]. Although this is just beginning work in utilizing the properties of spline functions in image restoration, the successful results indicate that further fruitful research can be performed in this field. Spline functions may be used in other areas of digital image processing such as image coding and reconstruction of images from their projections.

APPENDIX A

DERIVATION OF THE SMOOTHING SPLINE FILTER

Condition (5.18) implies $c_1 = c_N = 0$. Substituting (5.17) in (5.16a), (5.16b), (5.16c) and (5.16d) yields

$$a_{i+1} = a_i + b_i h_i + c_i h_i^2 + d_i h_i^3 \qquad i = 1, \ldots, N-1 \qquad (A-1)$$

$$b_{i+1} = b_i + 2c_i h_i + 3d_i h_i^2 \qquad i = 1, \ldots, N-1 \qquad (A-2)$$

$$c_{i+1} = c_i + 3d_i h_i \qquad i = 1, \ldots, N-1 \qquad (A-3)$$

$$3(d_i - d_{i+1}) = p(a_i - g_i)/\delta_i^2 \qquad i = 1, \ldots, N-1 \qquad (A-4)$$

where $h_i = x_{i+1} - x_i$, $i = 1, \ldots, N-1$ and $g_i = g(x_i)$. Obtaining d_i from (A-3), b_i from (A-1) and substituting these two values in (A-2) and (A-4) gives

$$c_i = (c_{i+1} - c_i)/(3h_i) \qquad (A-5)$$

$$b_i = (a_{i+1} - a_i)/h_i - c_i h_i - d_i h_i^2 \qquad (A-6)$$

$$\frac{1}{h_{i+1}} a_{i+2} - \left(\frac{1}{h_{i+1}} + \frac{1}{h_i}\right) a_{i+1} + \frac{1}{h_i} a_i = \frac{h_{i+1}}{3} c_{i+2} + 2\left(\frac{h_{i+1} + h_i}{3}\right) c_{i+1} + \frac{h_i}{3} c_i \qquad (A-7)$$

$$\frac{1}{h_{i+1}} c_{i+2} - \left(\frac{1}{h_{i+1}} + \frac{1}{h_i}\right) c_{i+1} + \frac{1}{h_i} c_i = p(a_i - g_i)/\delta_i^2 \,. \qquad (A-8)$$

Using vector space notation, Eqs. (A-7) and (A-8) can be written as

$$\underline{\Gamma}\,\underline{c} = \underline{Q}^t\underline{a} \tag{A-9}$$

$$\underline{Q}\,\underline{c} = p\underline{D}^{-2}(\underline{g}-\underline{a}) \tag{A-10}$$

where $\underline{c} = [c_2, \ldots, c_{N-1}]^t$, $\underline{a} = [a_1, \ldots, a_N]^t$, $\underline{\Gamma}$ is a positive definite tridiagonal matrix of order N-2 with elements

$$t_{ii} = 2\left(\frac{h_{i+1}+h_i}{3}\right), \quad t_{i,i+1} = t_{i+1,i} = \frac{h_{i+1}}{3}$$

and $\underline{Q}$ is a N x (N-2) tridiagonal matrix with elements

$$q_{i,i+1} = \frac{1}{h_i}, \quad q_{i+1,i+1} = -\frac{1}{h_{i+1}} - \frac{1}{h_i}, \quad q_{i+2,i+1} = \frac{1}{h_{i+1}} .$$

Premultiplying both sides of Eq. (A-10) by $\underline{Q}^t\underline{D}^2$

$$\underline{Q}^t\underline{D}^2\underline{Q}\,\underline{c} = p\underline{Q}^t\underline{g} - p\,\underline{Q}^t\underline{a} \tag{A-11}$$

and substituting $\underline{Q}^t\underline{a}$ from (A-9) into (A-11) gives

$$(\underline{Q}^t\underline{D}^2\underline{Q} + p\underline{\Gamma})\underline{c} = p\underline{Q}^t\underline{g} . \tag{A-12}$$

Since $\underline{Q}^t\underline{D}^2\underline{Q} + p\underline{\Gamma}$ is positive definite, $\underline{c}$ amy be obtained as

$$\underline{c} = p(\underline{Q}^t\underline{D}^2\underline{Q} + p\underline{\Gamma})^{-1}\underline{Q}^t\underline{g} . \tag{A-14}$$

From (A-10), $\underline{a}$ is derived as

$$\underline{a} = \underline{g} - p^{-1}\underline{D}^2\underline{Q}\,\underline{c} \tag{A-14}$$

Substituting (A-13) in (A-14) gives

$$\underline{a} = \underline{g} - \underline{D}^2\underline{Q}(\underline{Q}^t\underline{D}^2\underline{Q} + p\underline{\Gamma})^{-1}\underline{Q}^t\underline{g} .$$

Since $\hat{\underline{f}} = \underline{a}$, the filter is given as in Eq. (5.19).

APPENDIX B

MATRIX IDENTITY

Lemma: For any nonsingular matrices $\underline{B}$ and $\underline{C}$, the following matrix identity

$$(\underline{C}^{-1}+\underline{A}\,\underline{B}^{-1}\underline{A}^{t})^{-1} = \underline{C} - \underline{C}\,\underline{A}(\underline{A}^{t}\underline{C}\,\underline{A}+\underline{B})^{-1}\underline{A}^{t}\underline{C}$$

is valid for any matrix $\underline{A}$ if $(\underline{C}^{-1}+\underline{A}\,\underline{B}^{-1}\underline{A}^{t})^{-1}$ and $(\underline{A}^{t}\underline{C}\,\underline{A}+\underline{B})^{-1}$ exist.

Proof: Assuming an auxiliary matrix $\underline{D}$ of the form

$$\underline{D} = \underline{C}^{-1} + \underline{A}\,\underline{B}^{-1}\underline{A}^{t} \tag{B-1}$$

and premultiplying both sides of Eq. (B-1) by $\underline{D}^{-1}$ gives

$$I = \underline{D}^{-1}\underline{C}^{-1} + \underline{D}^{-1}\underline{A}\,\underline{B}^{-1}\underline{A}^{t}. \tag{B-2}$$

Postmultiplying (B-2) by $\underline{C}$, the equation

$$\underline{C} = \underline{D}^{-1} + \underline{D}^{-1}\underline{A}\,\underline{B}^{-1}\underline{A}^{t}\underline{C} \tag{B-3}$$

is obtained, and this is rewritten as

$$\underline{D}^{-1} = \underline{C} - \underline{D}^{-1}\underline{A}\,\underline{B}^{-1}\underline{A}^{t}\underline{C}\ . \tag{B-4}$$

Postmultiplying (B-3) by $\underline{A}$ gives

$$\begin{aligned}\underline{C}\,\underline{A} &= \underline{D}^{-1}\underline{A} + \underline{D}^{-1}\underline{A}\,\underline{B}^{-1}\underline{A}^{t}\underline{C}\,\underline{A} \\ &= \underline{D}^{-1}\underline{A}(I + \underline{B}^{-1}\underline{A}^{t}\underline{C}\,\underline{A}) \\ &= \underline{D}^{-1}\underline{A}\,\underline{B}^{-1}(\underline{B}+\underline{A}^{t}\underline{C}\,\underline{A})\ .\end{aligned} \tag{B-5}$$

Since the inverse of $\underline{B} + \underline{A}^t \underline{C}\ \underline{A}$ exists, $\underline{D}^{-1} \underline{A}\ \underline{B}^{-1}$ is obtained as

$$\underline{D}^{-1} \underline{A}\ \underline{B} = \underline{C}\ \underline{A}(\underline{B} + \underline{A}^t \underline{C}\ \underline{A})^{-1} \qquad \text{(B-6)}$$

and this can be substituted in (B-4) to give

$$\underline{D}^{-1} = \underline{C} - \underline{C}\ \underline{A}(\underline{B} + \underline{A}^t \underline{C}\ \underline{A})^{-1} \underline{A}^t \underline{C} \qquad \text{(B-7)}$$

or from (B-1)

$$(\underline{C}^{-1} + \underline{A}\ \underline{B}^{-1} \underline{A})^{-1} = \underline{C} - \underline{C}\ \underline{A}(\underline{B} + \underline{A}^t \underline{C}\ \underline{A})^{-1} \underline{A}^t \underline{C}\ . \qquad \text{(B-8)}$$

Q.E.D.

REFERENCES

1. H. C. Andrews, "Digital Image Restoration: A Survey," IEEE Computer, Vol. 7, No. 5, pp. 36-45, May 1974.

2. H. C. Andrews and C. L. Patterson, "Outer Product Expansions and Their Uses in Digital Image Processing," Aerospace Corp., El Segundo, Calif., Rep. ATR-74 (8139)-2, Jan. 1974.

3. A. Marechal, P. Croce and K. Dietzel, "Amelioration du Contraste de D'tails des Images Photographiques Par Filtrage des Frequences Spatiales," Opt. Acta, Vol. 5, pp. 256-262, 1958.

4. M. M. Sondhi, "Image Restoration: The Removal of Spatially Invariant Degradations," Proceedings of the IEEE, Vol. 60, No. 7, July 1972.

5. J. Tsujiuchi, "Corrections of Optical Images by Compensation of Aberrations and by Spatial Frequency Filtering," in Progress in Optics, Vol. 2, Wiley, New York, pp. 131-180, 1963.

6. J. L. Harris, Sr., "Image Evaluation and Restoration," J. Opt. Soc. Amer., Vol. 56, pp. 569-574, May 1974.

7. B. L. McGlamery. "Restoration of Turbulence-Degraded Images," J. Opt. Soc. Amer., Vol. 57, pp. 293-297, March 1967.

8. P. F. Mueller and G. O. Reynolds, "Image Restoration by Removal of Random Media Degradations," J. Opt. Soc. Amer., Vol. 57, pp. 1338-1344, November 1967.

9. C. W. Helstrom, "Image Restoration by the Method of Least Squares," J. Opt. Soc. Amer., Vol. 57, pp. 297-303, March 1967.

10. D. Slepian, "Linear Least-Squares Filtering of Distorted Images," J. Opt. Soc. Amer., Vol. 57, pp. 918-922, July 1967.

11. W. K. Pratt, "Generalized Wiener Filter Computation Techniques," IEEE Trans. Computers, pp. 636-641, July 1972.

12. A. Habibi, "Fast Suboptimal Wiener Filtering of Markov Processes," University of Southern California, Image Processing Institute Technical Report USCIPI Report 530, Los Angeles, CA 90007, pp. 75-80, March 1974.

13. F. Davarian, "Fast Computational Techniques for Pseudoinverse and Wiener Image Restoration," Ph.D. Dissertation, Department of Electrical Engineering, University of Southern California, Image Processing Institute, Technical Report USCIPI Report 610, Los Angeles, CA, 90007, August, 1975.

14. B. R. Hunt, "The Application of Constrained Least Squares Estimation to Image Restoration by Digital Computer," IEEE Transaction on Computers, Vol. C-22, No. 9, pp. 805-812, September 1973.

15. E. R. Cole, "The Removal of Unknown Image Blurs by Homomorphic Filtering," Ph.D. Dissertation, Department of Electrical Engineering, University of Utah, Salt Lake City, Utah, 84112, June 1973.

16. N. Mascarenhas, "Digital Image Restoration Under a Regression Model - The Unconstrained, Linear Equality and Inequality Constraint Approaches," Ph.D. Dissertation, Department of Electrical Engineering, University of Southern California, Image Processing Institute Technical Report, USCIPI 520, Los Angeles, CA, 90007, January 1974.

17. H. C. Andrews, "Positive Digital Image Restoration Techniques; A Survey," Aerospace Corp., El Segundo, Calif., Rep. ATR-73 (8138)-2, Feb. 1973.

18. A. V. Oppenheim, R. W. Schaefer, and T. G. Stockham, "Nonlinear Filtering of Multiplied and Convolved Signals," Proc. IEEE, Vol. 56, pp. 1264-1291, Aug. 1968.

19. T. G. Stockham, T. M. Cannon, R. B. Ingebretsen, "Blind Deconvolution Through Digital Signal Processing," Proc. IEEE, Vol. 63, No. 4, pp. 678-692, April 1975.

20. B. R. Hunt, "Digital Signal Processing," Proc. IEEE, Vol. 63, No. 4, pp. 693-708, April 1975.

21. L. J. Cutrona, "Recent Development in Coherent Optical Techniques," in Optical and Electro-Optical Information Processing, J. T. Tippett etal., Eds., M.I.T. Press, Cambridge, Mass. 1965.

22. A. A. Sawchuk, "Space-Variant Image Restoration by Coordinate Transformation," J. Opt. Soc. Amer., Vol. 64, No. 2, pp. 138-144, Feb. 1974.

23. J. W. Goodman, Introduction to Fourier Optics, McGraw-Hill, New York, 1968.

24. A. A. Sawchuk, and M. J. Peyrovian, "Restoration of Astigmatism and Curvature of Field," J. Opt. Soc. Amer., Vol. 65, No. 4, pp. 712-715, June 1975.

25. A. A. Sawchuk, "Space-Variant Image Motion Degradation and Restoration," Proceedings of IEEE, Vol. 60, No. 7, July 1972.

26. G. M. Robbins and T. S. Huang, "Inverse Filtering for Linear Shift-Variant Imaging Systems," Proceedings of IEEE, vol. 60, No. 7, July 1972.

27. T. S. Huang, "Some Notes on Film Grain Noise," in Woods Holes Summer Study Rep. on Restoration of Atmospherically Degraded Images, Vol. 2, Alexandria, VA., Defense Documentation Center, pp. 105-109, July 1966.

28. F. Naderi and A. A. Sawchuk, "Nonlinear Detection and Estimation of Images Degraded by Film-Grain Noise," Proc. Opt. Soc. Amer., Topical Meeting on Image Processing, Asilomar, Calif., Feb. 24-26, 1976, pp. 17-24.

29. A. G. Tescher and H. C. Andrews, "Data Compression and Enhancement of Sampled Images," Applied Optics, Vol. 13, No. 1, January, 1974.

30. T. N. Cornsweet, Visual Perception, Academic Press, New York, 1970.

31. I. J. Schoenberg, "Monospline and Quadrature Formulae," Theory and Application of Spline Functions, T. N. E. Greville, pp. 157-207, Academic Press, New York and London, 1969.

32. A. Ralston, A First Course in Numerical Analysis, McGraw-Hill, New York, 1965.

33. T. N. E. Greville, Theory and Application of Spline Functions, Academic Press, New York, 1969.

34. J. C. Holladay, "Smoothest Curve Approximation," Math. Tables Aids Comput. 11, pp. 233-243, 1957.

35. C. de Boor, "Best Approximation Properties of Spline Functions of Odd Degree," J. Math. Mech. 12, pp. 747-749, 1963.

36. I. J. Schoenberg, "Spline Functions and the Problems of Graduation," Proc. Nat. Acad. Sci. U.S.A. 52, pp. 947-950, 1964.

37. L. L. Schumaker, "Some Algorithms for the Computation of Interpolating and Approximating Spline Functions," Theory and Application of Spline Functions, T. N. E. Greville, pp. 87-102, Academic Press, New York, 1969.

38. H. B. Curry and I. J. Schoenberg, "On Polya Frequency Functions IV. The Fundamental Spline Functions and Their Limits," J. Analyse Math. 17, pp. 71-107, 1966.

39. M. H. Schultz, Spline Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1973.

40. I. J. Schoenberg, "On Monospline of Least Square Deviation and Best Quadrature Formulae," SIAM J. Numer Anal. Ser. B-2, pp. 144-170, 1965.

41. I. J. Schoenberg, "On Monospline of Least Square Deviation and Best Quadrature Formulae II," SIAM J. Numer Anal. 3, pp. 321-328, 1966.

42. F. Davarian and M. J. Peyrovian, "Quadrature Formulae Application in Image Restoration," University of Southern California, Image Processing Institute Technical Report, USCIPI Report 560, Los Angeles, CA, 90007, pp. 60-67, March 1975.

43. M. J. Peyrovian and A. A. Sawchuk, "Image Restoration by Smoothing Spline Functions," University of Southern California, Image Processing Institute Technical Report, USCIPI Report 620, pp. 86-92, September 1975.

44. A. Albert, Regression and the Moore-Penrose Pseudoinverse, Academic Press, New York, 1972.

45. H. S. Hou, "Least Squares Image Restoration Using Spline Interpolation," Ph.D. Dissertation, Department of Electrical Engineering, University of Southern California, Image Processing Institute, Technical Report USCIPI Report 650, pp. 86-92, September 1975.

46. G.H. Golub and C. Reinsch, "Singular Value Decomposition and Least Squares Solutions," Numer. Math., Vol. 14, pp. 403-420, 1970.

47. M. Born and E. Wolf, Principles of Optics, Fourth Edition, Pergamon Press, London, 1970.

48. G. M. Robbins, Mass. Inst. Tech Res. Lab Electronics Quart. Prog. Report, No. 93, pp. 243-251, 1969.

49. W. K. Pratt, Digital Image Processing, to be published.

50. A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York, 1965.

51. P. Liebelt, An Introduction to Optimal Estimation, Addison-Wesley, Menlo Park, Calif, 1969.

52. B. R. Hunt and H. C. Andrews, "Comparison of Different Filter Structures for Image Restoration," Proc. 6th Annual Hawaii Int. Conf. on System Sciences, Jan. 1973.

53. C. H. Reinsch, "Smoothing by Spline Functions," Numer. Math., Vol. 10, pp. 177-183, 1967.

54. C. H. Reinsch, "Smoothing by Spline Functions, II" Numer. Math., Vol. 16, pp. 451-454, 1971.

55. R. S. Martin and J. H. Wilkinson, "Symmetric Decomposition of Positive Definite Banded Matrices," Numer. Math., vol. 7, pp. 355, 1965.

56. J. F. Walkup and R. C. Choens, "Image Processing in Signal-Dependent Noise," Optical Engineering, Vol. 13, pp. 258-266, 1974.

57. E. L. Allgower, "Exact Inverses of Certain Band Matrices," Numer. Math., Vol. 21, pp. 279-284, 1973.

58. F. Naderi, "Estimation and Detection of Images Degraded by Film-Grain Noise," Ph.D. Dissertation, Department of Electrical Engineering, University of Southern California, Image Processing Institute, Los Angeles, CA, 90007, August, 1976.

59. B. R. Hunt and T. M. Cannon, "Nonstationary Assumptions for Gaussian Models of Images," To appear in IEEE Trans. Systems, Man and Cybernetics.

60. C. de Boor and J. R. Rice, "Least Square Cubic Spline Approximation I: Fixed Knots," Computer Science Department, Purdue University, Rep. CSD TR 20, Lafayette, Indiana, 47907, 1968.

61. _____, "Least Square Cubic Spline Approximation II: Variable Knots," Computer Science Department, Purdue University, Rep. CDS TR 21, Lafayette, Indiana, 47907, 1968.

62. T. S. Huang, D. S. Baker, and S. P. Berger, "Iterative Image Restoration," Applied Optics, Vol. 14, No. 5, pp. 1165-1168, May 1975.

63. T. S. Huang, "Restoring Images with Shift-Varying Degradations," Proc. Opt. Soc. Amer. Toptical Meeting on Image Processing, Asilomar, California, pp. 149-151, Feb. 24-26, 1976.

64. H. C. Andrews, and C. L. Patterson, "Digital Interpolation of Discrete Images," IEEE Trans. Computer, pp. 196-202, February, 1976.